Target genes, variants, tissues and transcriptional pathways for the regulation of serum urate levels in humans

Adrienne Tin*†[1,2], Jonathan Marten*[3], Victoria L. Halperin Kuhns *[4], Yong Li*[5], Matthias Wuttke*[5], Holger Kirsten*[6,7], Karsten B. Sieber[8], Chengxiang Qiu[9], Mathias Gorski[10,11], Zhi Yu[1,12], Ayush Giri[13,14], Gardar Sveinbjornsson[15], Man Li[16], Audrey Y. Chu[17], Anselm Hoppmann[5], Luke J. O'Connor[18], Bram Prins[19], Teresa Nutile[20], Damia Noce[21], Masato Akiyama[22,23], Massimiliano Cocca[24], Sahar Ghasemi[25,26], Peter J. van der Most[27], Katrin Horn[6,7], Yizhe Xu[16], Christian Fuchsberger[21], Sanaz Sedaghat[28], Saima Afaq[29,30], Najaf Amin[28], Johan Ärnlöv[31,32], Stephan J.L. Bakker[33], Nisha Bansal[34,35], Daniela Baptista[36], Sven Bergmann[37,38,39], Mary L. Biggs[40,41], Ginevra Biino[42], Eric Boerwinkle[43], Erwin P. Bottinger[44,45], Thibaud S. Boutin[3], Marco Brumat[46], Ralph Burkhardt[7,47,48], Eric Campana[46], Archie Campbell[49], Harry Campbell[50], Robert J. Carroll[51], Eulalia Catamo[24], John C. Chambers[52,53,54,55,56], Marina Ciullo[20,57], Maria Pina Concas[24], Josef Coresh[1], Tanguy Corre[37,38,58], Daniele Cusi[59,60], Sala Cinzia Felicita[61], Martin H. de Borst[33], Alessandro De Grandi[21], Renée de Mutsert[62], Aiko P.J. de Vries[63], Graciela Delgado[64], Ayse Demirkan[28], Olivier Devuyst[65], Katalin Dittrich[66,67], Kai-Uwe Eckardt[68,69], Georg Ehret[36], Karlhans Endlich[26,70], Michele K. Evans[71], Ron T. Gansevoort[33], Paolo Gasparini[24,46], Vilmantas Giedraitis[72], Christian Gieger[73,74,75], Giorgia Girotto[24,46], Martin Gögele[21], Scott D. Gordon[76], Daniel F. Gudbjartsson[15], Vilmundur Gudnason[77,78], Toomas Haller[79], Pavel Hamet[80,81], Tamara B. Harris[82], Caroline Hayward[3], Andrew A. Hicks[21], Edith Hofer[83,84], Hilma Holm[15], Wei Huang[85,86], Nina Hutri-Kähönen[87,88], Shih-Jen Hwang[89,90], M. Arfan Ikram[28], Raychel M. Lewis[91], Erik Ingelsson[92,93,94,95], Johanna Jakobsdottir[77,96], Ingileif Jonsdottir[15], Helgi Jonsson[97,98], Peter K. Joshi[50], Navya Shilpa Josyula[99], Bettina Jung[10], Mika Kähönen[100,101], Yoichiro Kamatani[22,102], Masahiro Kanai[22,103], Shona M. Kerr[3], Wieland Kiess[7,66,67], Marcus E. Kleber[64], Wolfgang Koenig[104,105,106], Jaspal S. Kooner[54,55,56,107], Antje Körner[7,66,67], Peter Kovacs[108], Bernhard K. Krämer[64], Florian Kronenberg[109], Michiaki Kubo[110], Brigitte Kühnel[73], Martina La Bianca[24], Leslie A. Lange[111], Benjamin Lehne[29], Terho Lehtimäki[112,113], Lifelines Cohort Study[114], Jun Liu[28], Markus Loeffler[6,7], Ruth J.F. Loos[44,115], Leo-Pekka Lyytikäinen[112,113], Reedik Magi[79], Anubha Mahajan[116,117], Nicholas G. Martin[76], Winfried März[64,118,119], Deborah Mascalzoni[21], Koichi Matsuda[120], Christa Meisinger[121,122], Thomas Meitinger[105,123,124], Andres Metspalu[79], Yuri Milaneschi[125], Million Veteran Program[126], Christopher J. O'Donnell[127,128], Otis D. Wilson[129], J. Michael Gaziano[130], Pashupati P. Mishra[131], Karen L. Mohlke[132], Nina Mononen[112,131], Grant W. Montgomery[133], Dennis O. Mook-Kanamori[62,134], Martina Müller-Nurasyid[105,135,136,137], Girish N. Nadkarni[44,138], Mike A. Nalls[139,140], Matthias Nauck[26,141], Kjell Nikus[142,143], Boting Ning[144], Ilja M. Nolte[27], Raymond Noordam[145], Jeffrey OConnell[146], Isleifur Olafsson[147], Sandosh Padmanabhan[148], Brenda W.J.H. Penninx[125], Thomas Perls[149], Annette Peters[74,75,105], Mario Pirastu[150], Nicola Pirastu[50], Giorgio Pistis[151], Ozren Polasek[152,153], Belen Ponte[154,155], David J. Porteous[49,156], Tanja Poulain[7], Michael H. Preuss[44], Ton J. Rabelink[63,157], Laura M. Raffield[132], Olli T. Raitakari[158,159], Rainer Rettig[160], Myriam Rheinberger[10], Kenneth M. Rice[41], Federica Rizzi[161,162], Antonietta Robino[24], Igor Rudan[50], Alena Krajcoviechova[163,164], Renata Cifkova[163,164], Rico Rueedi[37,38], Daniela Ruggiero[20,57], Kathleen A. Ryan[165], Yasaman Saba[166], Erika Salvi[161,167], Helena Schmidt[168], Reinhold Schmidt[83], Christian M. Shaffer[51], Albert V. Smith[78], Blair H. Smith[169], Cassandra N. Spracklen[132], Konstantin Strauch[135,136], Michael Stumvoll[170], Patrick Sulem[15], Salman M. Tajuddin[71], Andrej Teren[7,171], Joachim Thiery[7,47], Chris H. L. Thio[27], Unnur Thorsteinsdottir[15], Daniela Toniolo[61], Anke Tönjes[172], Johanne Tremblay[80,173], André G. Uitterlinden[174], Simona Vaccargiu[150], Pim van der Harst[175,176,177], Cornelia M. van Duijn[28], Niek Verweij[175], Uwe Völker[26,178], Peter Vollenweider[179], Gerard Waeber[179], Melanie Waldenberger[73,74,105], John B. Whitfield[76], Sarah H. Wild[180], James F. Wilson[3,50], Qiong Yang[144], Weihua Zhang[52,54], Alan B. Zonderman[71], Murielle Bochud[58], James G. Wilson[181], Sarah A. Pendergrass[182], Kevin Ho[183,184], Afshin Parsa[185,186], Peter P. Pramstaller[21], Bruce M. Psaty[187,188], Carsten A. Böger[10,189], Harold Snieder[27], Adam S. Butterworth[190], Yukinori Okada[191,192], Todd L. Edwards[193,194], Kari Stefansson[15], Katalin Susztak[9], Markus Scholz[6,7], Iris M.

Heid[11], Adriana M. Hung**[129,194], Alexander Teumer**[25,26], Cristian Pattaro**[21], Owen M. Woodward**[4], Veronique Vitart**[3], Anna Köttgen**†[1,5]

* Indicates joint contribution
** Indicates joint oversight
† Indicates corresponding author

Authors for Correspondence:

Adrienne Tin, PhD MS
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
Baltimore, Maryland, USA
+1 443-287-4740
atin1@jhu.edu

Anna Köttgen, MD MPH
Institute of Genetic Epidemiology
Medical Center - University of Freiburg
Hugstetter Str. 49, 79106 Freiburg, Germany
+49 (0)761 270-78050
anna.koettgen@uniklinik-freiburg.de

Author affiliations

1 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
2 Welch Centre for Prevention, Epidemiology and Clinical Research, Baltimore, Maryland, USA
3 Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
4 Department of Physiology, University of Maryland School of Medicine, Baltimore MD, USA
5 Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, Germany
6 Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany
7 LIFE Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany
8 Target Sciences - Genetics, GlaxoSmithKline, Collegeville, Pennsylvania, USA
9 Smilow Center for Translational Research, Perelman School of Medicine, University of Pennsylvania
10 Department of Nephrology, University Hospital Regensburg, Regensburg, Germany
11 Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany
12 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
13 Division of Quantitative Sciences, Department of Obstetrics & Gynecology, Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA

14 Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA
15 deCODE Genetics, Amgen Inc., Reykjavik, Iceland
16 Department of Medicine, Division of Nephrology and Hypertension, University of Utah, Salt Lake City, USA
17 Genetics, Merck & Co., Inc., Kenilworth, New Jersey, USA
18 Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
19 Strangeways Research Laboratory, University of Cambridge, 2 Worts' Causeway, Cambridge, CB1 8RN, UK
20 Institute of Genetics and Biophysics Adriano Buzzati-Traverso - CNR, Naples, Italy
21 Eurac Research, Institute for Biomedicine (affiliated to the University of Lübeck), Bolzano, Italy
22 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan
23 Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
24 Institute for Maternal and Child Health - IRCCS Burlo Garofolo, Trieste, Italy
25 Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
26 DZHK (German Center for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany
27 Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
28 Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands
29 Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial College London, London, UK
30 Institute of Public health & social sciences, Khyber Medical University, Pakistan
31 Department of Neurobiology, Care Sciences and Society, Division of Family Medicine and Primary Care, Karolinska Institutet, Stockholm, Sweden
32 School of Health and Social Studies, Dalarna University, Sweden
33 Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
34 Division of Nephrology, University of Washington, Seattle, Washington, USA
35 Kidney Research Institute, University of Washington, Seattle, Washington, USA
36 Cardiology, Geneva University Hospitals, Geneva, Switzerland
37 Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
38 Swiss Institute of Bioinformatics, Lausanne, Switzerland
39 Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa
40 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA
41 Department of Biostatistics, University of Washington, Seattle, Washington, USA
42 Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy
43 Human Genetics Centre, University of Texas Health Science Centre, Houston, Texas, USA
44 The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA
45 Digital Health Centre, Hasso Plattner Institute and University of Potsdam, Potsdam, Germany
46 University of Trieste, Department of Medicine, Surgery and Health Sciences, Trieste, Italy
47 Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig, Leipzig, Germany

48 Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany
49 Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
50 Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
51 Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville (Tennessee), USA
52 Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial
113 College London, London, UK
53 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore
54 Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, UK
55 Imperial College Healthcare NHS Trust, Imperial College London, London, UK
56 MRC-PHE Centre for Environment and Health, 323 School of Public Health, Imperial College London, London, UK
57 IRCCS Neuromed, Pozzilli, Italy
58 Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland
59 Institute of Biomedical Technologies, Italy National Research Council, Segrate (Milano), Italy
60 Bio4Dreams - business nursery for life sciences, Bresso (Milano), Italy
61 San Raffaele Research Institute, Milano, Italy
62 Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands
63 Section of Nephrology, Department of Internal Medicine, Leiden University Medical Centre, Leiden, The Netherlands
64 5th Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
65 Institute of Physiology, University of Zurich, Zurich, Switzerland
66 Department of Women and Child Health, Hospital for Children and Adolescents, University of Leipzig, Leipzig, Germany
67 Centre for Pediatric Research, University of Leipzig, Leipzig, Germany
68 Intensive Care Medicine, Charité, Berlin, Germany
69 Department of Nephrology and Hypertension, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany
70 Department of Anatomy and Cell Biology, University Medicine Greifswald, Greifswald, Germany
71 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Baltimore, Maryland, USA
72 Department of Public Health and Caring Sciences, Molecular Geriatrics, Uppsala University, Uppsala, Sweden
73 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
74 Institute of Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
75 German Center for Diabetes Research (DZD), Neuherberg, Germany
76 QIMR Berghofer Medical Research Institute, Brisbane, Australia
77 Icelandic Heart Association, Kopavogur, Iceland
78 Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland
79 Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
80 Montreal University Hospital Research Centre, CHUM, Montreal, Canada

81 Medpharmgene, Montreal, Canada
82 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda, Maryland, USA
83 Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria
84 Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria
85 Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Centre, Shanghai, China
86 Shanghai Industrial Technology Institute, Shanghai, China
87 Department of Pediatrics, Tampere University Hospital, Tampere, Finland
88 Department of Pediatrics, Faculty of Medicine and Life Sciences, University of Tampere, Finland
89 NHLBIs Framingham Heart Study, Framingham (Massachusetts), USA
90 The Centre for Population Studies, NHLBI, Framingham (Massachusetts), USA
91 Department of Physiology, University of Maryland School of Medicine
92 Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, USA
93 Stanford Cardiovascular Institute, Stanford University, USA
94 Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
95 Stanford Diabetes Research Center, Stanford University, Stanford, USAs
96 The Centre of Public Health Sciences, University of Iceland, Reykjavik, Iceland
97 Landspitalinn University Hospital, Iceland
98 University of Iceland, Iceland
99 Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, USA
100 Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland
101 Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
102 Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
103 Department of Biomedical Informatics, Harvard Medical School, Boston, USA
104 Deutsches Herzzentrum München, Technische Universität München, Munich, Germany
105 DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany
106 Institute of Epidemiology and Biostatistics, University of Ulm, Ulm, Germany
107 National Heart and Lung Institute, Imperial College London, London W12 0NN, UK
108 Integrated Research and Treatment Centre Adiposity Diseases, University of Leipzig, Leipzig, Germany
109 Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria
110 RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan
111 Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of Colorado Denver - Anschutz Medical Campus, Aurora (Colorado), USA
112 Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland
113 Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, Tampere University, Tampere, Finland
114 Lifelines Cohort Study, Groningen, the Netherlands

115 The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA
116 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
117 Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, UK
118 Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany
119 Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Austria
120 Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
121 Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany
122 Chair of Epidemiology Ludwig- Maximilians-Universität München at UNIKA-T Augsburg, Augsburg, Germany
123 Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
124 Institute of Human Genetics, Technische Universität München, Munich, Germany
125 Department of Psychiatry, Amsterdam Neuroscience and Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, Amsterdam, The Netherlands
126 Department of Veterans Affairs, Office of Research and Development, Washington, DC, USA
127 VA Boston Healthcare System, Boston, MA, USA
128 Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
129 Vanderbilt University Medical Centre, Division of Nephrology & Hypertension, Nashville, TN, USA
130 Massachusetts Veterans Epidemiology Research and Information Center, VA Cooperative Studies Program, VA Boston Healthcare System, Boston (Massachusetts), USA
131 Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
132 Department of Genetics, University of North Carolina, Chapel Hill (North Carolina), USA
133 University of Queensland, St Lucia, Australia
134 Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, The Netherlands
135 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
136 Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Germany
137 Department of Internal Medicine I (Cardiology), Hospital of the Ludwig-Maximilians-University (LMU) Munich, Munich, Germany
138 Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA
139 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA
140 Data Tecnica International, Glen Echo, Maryland, USA
141 Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
142 Department of Cardiology, Heart Center, Tampere University Hospital, Tampere, Finland
143 Department of Cardiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, Tampere University, Tampere, Finland
144 Department of Biostatistics, Boston University School of Public Health, Boston (Massachusetts), USA
145 Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Centre, Leiden, The Netherlands

146 University of Maryland School of Medicine, Baltimore, USA
147 Department of Clinical Biochemistry, Landspitali University Hospital, Reykjavik, Iceland
148 Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK
149 Department of Medicine, Geriatrics Section, Boston Medical Center, Boston University School of Medicine, Boston (Massachusetts), USA
150 Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Li Punti (Sassari), Italy
151 Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland
152 Faculty of Medicine, University of Split, Split, Croatia
153 Gen-info Ltd, Zagreb, Croatia
154 Service de Néphrologie, Geneva University Hospitals, Geneva, Switzerland
155 Nephrology Service, Department of Specialties in Internal Medicine, University Hospitals of Geneva, Switzerland
156 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
157 Einthoven Laboratory of Experimental Vascular Research, Leiden University Medical Centre, Leiden, The Netherlands
158 Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
159 Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
160 Institute of Physiology, University Medicine Greifswald, Karlsburg, Germany
161 Department of Health Sciences, University of Milan, Milano, Italy
162 ePhood Scientific Unit, ePhood SRL, Milano, Italy
163 Centre for Cardiovascular Prevention, First Faculty of Medicine, Department of Medicine, Charles University in Prague, Prague, Czech Republic
164 Thomayer Hospital, Prague, Czech Republic
165 Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, USA
166 Molecular Biology and Biochemistry, Gottfried Schatz Research Centre for Cell Signaling, Metabolism and Aging, Medical University of Graz, Graz, Austria
167 Neuroalgology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
168 Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz, Austria
169 Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK
170 Division of Endocrinology, Nephrology and Rheumatology, University of Leipzig, Leipzig, Germany
171 Heart Centre Leipzig, Leipzig, Germany
172 Department of Endocrinology and Nephrology, University of Leipzig, Leipzig, Germany
173 CRCHUM, Montreal, Canada
174 Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands
175 Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
176 Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
177 Durrer Centre for Cardiovascular Research, The Netherlands Heart Institute, Utrecht, The Netherlands

178 Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
179 Internal Medicine, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland
180 Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
181 Department of Physiology and Biophysics, University of Mississippi Medical Centre, Jackson (Mississippi), USA
182 Geisinger Research, Biomedical and Translational Informatics Institute, Danville, Pennsylvania, USA
183 Kidney Health Research Institute (KHRI), Geisinger, Danville, Pennsylvania, USA
184 Department of Nephrology, Geisinger, Danville, Pennsylvania, USA
185 Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, USA
186 Department of Medicine, University of Maryland School of Medicine, Baltimore, USA
187 Cardiovascular Health Research Unit, Department of Medicine, Department of Epidemiology, Department of Health Service, University of Washington, Seattle, Washington, USA
188 Kaiser Permanente Washington Health Research Institute, Seattle, Washington, USA
189 Department of Nephrology and Rheumatology, Kliniken Südostbayern AG, Regensburg, Germany
190 Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
191 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Osaka, Japan
192 Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan
193 Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville, TN, USA
194 Department of Veterans Affairs, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA

Abstract

Elevated serum urate levels cause gout, and correlate with cardio-metabolic diseases via poorly understood mechanisms. We performed a trans-ethnic genome-wide association study of serum urate among 457,690 individuals, identifying 183 loci (147 novel) that improve prediction of gout in an independent cohort of 334,880 individuals. Serum urate showed significant genetic correlations with many cardio-metabolic traits, with genetic causality analyses supporting a substantial role for pleiotropy. Enrichment analysis, fine-mapping of urate-associated loci and co-localization with gene expression in 47 tissues implicated kidney and liver as main target organs and prioritized potentially causal genes and variants, including the transcriptional master regulators in liver and kidney, *HNF1A* and *HNF4A*. Experimental validation showed that *HNF4A* trans-activated the promoter of the major urate transporter *ABCG2* in kidney cells, and that HNF4A p.Thr139Ile is a functional variant. Transcriptional co-regulation within and across organs may be a general mechanism underlying the observed pleiotropy between urate and cardio-metabolic traits.

Introduction

Serum urate levels reflect a balance between uric acid production and its net excretion via kidney and intestine. Elevated serum urate levels define hyperuricemia, which is associated with components of the metabolic syndrome as well as with cardiovascular and kidney disease. Hyperuricemia can cause kidney stones and gout, the most common form of inflammatory arthritis[1,2]. Gout attacks are a highly painful inflammatory response to the deposition of urate crystals, and are a significant cause of morbidity, emergency room visits and related health care costs[3]. Although gout has become a major public health issue, it is undertreated due to low awareness, poor patient adherence[4] and inappropriate prescription practices of the most commonly used drug, allopurinol[5]. A better understanding of the mechanisms controlling serum urate levels may not only help to develop novel medications to treat and prevent gout, but also provide insights into regulatory mechanisms shared with urate-associated cardio-metabolic risk factors and diseases.

Genetic heritability of serum urate varies between 30% and 60% in diverse populations[6-11]. Candidate gene and early genome-wide associations studies (GWAS) have identified three genes as major determinants of urate levels: *SLC2A9*, *ABCG2*, and *SLC22A12*[7,12-18]. While *SLC2A9* and *ABCG2* harbor common variants of relatively large effect[19], *SLC22A12* contains many rare or low-frequency variants associated with lower serum urate levels[20]. The largest GWAS meta-analyses performed to date identified 28 associated genomic loci among European ancestry (EA) individuals[21] and 27 among Japanese individuals[22]. Many genes in the associated loci encode urate transporters or their regulators in kidney and gut, while others are relevant to glucose and lipid metabolism, central functions of the liver, where uric acid is generated. Earlier GWAS did not perform fine-mapping coupled to functional annotation or co-localization with gene expression across tissues to prioritize target tissues, pathways, and potentially causal genes and variants. These approaches have only recently become available owing to increased public availability of large datasets[23,24].

Here, we perform a trans-ethnic GWAS meta-analysis of serum urate among 457,690 individuals and identify 183 associated genetic loci that improve risk prediction of gout in an independent sample of 334,880 individuals from the UK Biobank. We evaluate the genetic correlation of serum urate with hundreds of cardio-metabolic traits and diseases, and use a recently developed latent causal variable model to examine the contribution of causality versus pleiotropy. We prioritize target variants, genes, tissues and pathways that contribute to the complex regulation of urate levels through comprehensive data integration. To validate the prioritization workflow, we conduct proof-of-principle experimental studies showing that HNF4A, a transcriptional master regulator in liver and kidney proximal tubule, can regulate transcription of the major urate transporter ABCG2 in kidney cells and that the fine-mapped *HNF4A* variant p.Thr139Ile is functional. Transcriptional co-regulation of processes linked to energy metabolism within and across organs may underlie the pleiotropy we uncovered between urate levels and numerous cardio-metabolic traits.

Results

Meta-analyses and characterization of serum urate-associated loci

Overview

Trans-ethnic meta-analyses were conducted to maximize the sample size for studying the genetic landscape of serum urate. EA-specific analyses were used where population-specific linkage disequilibrium (LD) was required for prioritizing urate-associated genes, tissues, and pathways, identifying genetic correlations with other traits, and to perform gout risk prediction (**Supplementary Figure 1**).

Trans-ethnic meta-analysis identifies 183 loci associated with serum urate

The primary trans-ethnic GWAS meta-analysis included 457,690 individuals (EA, n=288,649; East Asian ancestry [EAS], n=125,725; African Americans [AA], n=33,671; South Asian ancestry [SA], n=9,037; and Hispanics [HIS], n=608) from 74 studies, with mean urate levels ranging from 4.2 to 7.2 mg/dl (**Supplementary Table 1**). GWAS were performed based on genotypes imputed

using the 1000 Genomes Project or Haplotype Reference Consortium reference panels (Methods, **Supplementary Table 2**). Following standardized study-specific quality control and variant filtering procedures, we combined results through inverse-variance weighted fixed effect meta-analysis. There was no evidence of inflation due to unmodeled population structure (LD score regression intercept=1.01; genomic inflation factor λ_{GC}=1.04). Post-meta-analysis variant filtering left 8,249,849 high-quality SNPs for downstream analyses (Methods).

We identified 183 loci, defined as the +/-500 kb region around the SNP with the lowest p-value (index SNP), that contained at least one SNP associated at genome-wide significance ($p \leq 5x10^{-8}$, **Figure 1**, **Supplementary Table 3**). Of these, 36 contained a SNP reported as the index SNP in previous GWAS of serum urate[13,15,17,18,21,22,25,26], and 147 were considered novel (**Figure 1**). Absolute effect estimates of each copy of the respective index SNP on serum urate ranged from 0.28 mg/dl (known *SLC2A9* locus) to 0.017 mg/dl (novel *KLB* locus). The average absolute effect across all index SNPs was of 0.038 mg/dl (standard deviation [SD] 0.033). Regional association plots for all 183 loci are shown in **Supplementary Figure 2**.

The index SNPs at all 183 loci explained an estimated 7.7% of the serum urate variance (Methods), as compared to 5.3% of the variance explained by variants previously reported by GWAS in EA populations[21]. In a large participating general population-based pedigree study, the 183 index SNPs explained 17% of serum urate genetic heritability (h^2=37%, 95% credible interval: 29%, 45%). The index SNPs at the three major urate loci *SLC2A9*, *ABCG2* and *SLC22A12* explained 5% of the genetic heritability (**Supplementary Figure 3**; Methods).

Characterization of ancestry-related heterogeneity

For the 183 index SNPs, we observed no evidence of systematic between-study heterogeneity (median I^2=2%, interquartile range 0-14%; **Supplementary Table 3**). Fourteen index SNPs showed significant evidence of ancestry-associated heterogeneity ($p_{anc\text{-}het} < 2.7x10^{-4} = 0.05/183$) when tested using meta-regression (**Supplementary Figure 4**, Methods), consistent with their higher measures of between-study heterogeneity (I^2>25%, **Figure 1**, **Supplementary Table 3**). The most significant ancestry-associated heterogeneity was observed for the index SNP rs3775947 at *SLC2A9* ($p_{anc\text{-}het}=1.5x10^{-127}$, effect per copy of the coded allele in EA 0.34 mg/dl, AA

0.26 mg/dl, EAS 0.17 mg/dl, HIS 0.41 mg/dl, and SA 0.21 mg/dl), consistent with previous reports of population heterogeneity of genetic effects at this locus[27]. In addition, nine genome-wide significant loci were identified through meta-regression that did not overlap with the 183 significant loci from the primary trans-ethnic fixed-effects meta-analysis. Of these, the index SNPs at *SLC2A2* and *KCNQ1* were genome-wide significant in EAS (**Supplementary Table 4**). Results from ancestry-specific meta-analyses of EA, AA, EAS and SA are summarized in **Supplementary Tables 5 to 8**, respectively, as well as in the **Supplementary Information**.

Sex-stratified meta-analyses of serum urate GWAS

Mean serum urate levels and gout risk are higher in men than in women[28]. We therefore performed secondary sex-specific analyses to evaluate whether the 183 urate-associated index SNPs showed sex-specific differences. After multiple-testing correction, six SNPs showed significant effect differences ($P_{diff}<2.7x10^{-4}=0.05/183$), at *SLC2A9*, *ABCG2*, *CAPN1*, *GCKR*, *IDH2*, and *SLC22A12* (**Supplementary Table 9**). A genome-wide test for differences in genetic effects on urate levels between men and women identified only SNPs at *SLC2A9* and *ABCG2* as significant ($p_{diff}<5x10^{-8}$, Methods*,* **Supplementary Figure 5**), consistent with previous reports,[7,14,15,21] with few additional loci outside of the extended sex-specific regions that were suggestive of sex differences ($p_{diff}<1x10^{-5}$, **Supplementary Table 10**).

Epidemiological and clinical landscape

Urate-associated SNPs are associated with gout

To assess the association of the 183 trans-ethnic urate index SNPs with gout, we investigated their effects in a trans-ethnic meta-analysis of gout from 20 studies, based on 763,813 participants including 13,179 with gout (Methods, **Figure 1**, **Supplementary Table 1**). Consistent with the causal role of hyperuricemia in gout, genetic effects were highly correlated (Spearman correlation coefficient 0.87, **Supplementary Figure 6A**); 55 SNPs were significantly associated with gout ($p<2.7x10^{-4}=0.05/183$). In agreement with previous findings[29], the largest odds ratio (OR) for gout was observed at *ABCG2* (rs74904971, OR 2.04, 95% confidence interval [CI] 1.96-

2.12, P=7.7x10^{-299}). The genetic effects were generally higher among index SNPs with lower minor allele frequency (MAF), with the exception of a few large-effect SNPs with MAF>10%, mapping into loci that encode urate transporters with known major effects on urate levels: *SLC2A9*, *ABCG2*, and *SLC22A12*[30] (**Supplementary Figure 6B**).

A genetic risk score for urate improves risk prediction for gout

We evaluated whether a weighted urate genetic risk score (GRS) improved risk prediction of gout when added to demographic information in a large, independent sample of 334,880 individuals from the UK Biobank (UKBB), including 4,908 gout cases (Methods). Across categories of the urate GRS, gout prevalence increased from 0.1% in the lowest GRS category to 12.9% in the highest GRS category (**Figure 2A, Supplementary Table 11**). Using the most common GRS category as the reference, the age- and sex-adjusted OR of gout ranged from 0.09 (95% CI 0.02-0.37, p=7.8x10^{-4}) in the lowest GRS category to 13.6 (95% CI 7.2-25.7, p=1.4x10^{-15}) in the highest GRS category (**Figure 2B, Supplementary Table 11**). Of note, the 3.5% of individuals in the three highest GRS categories had a >3-fold increase in the risk of gout compared to individuals in the most common GRS category. This risk is comparable to a monogenic disease of modest effect size[31], but affects a comparatively higher proportion of the population.

We additionally constructed gout risk prediction models in the independent UK Biobank sample by regressing gout status on the GRS alone ("genetic model"), on age and sex ("demographic model"), and on the GRS, age, and sex ("combined model") in a random training subset consisting of 90% of the individuals. These models were then used to predict gout status in the remaining 10%. The genetic model (area under the receiver operating characteristic curve [AUC]=0.68) was a weaker predictor than the demographic model (AUC=0.79), but addition of the GRS to the demographic model (combined model) significantly increased the prediction accuracy (AUC=0.84, DeLong's test p<2.2x10^{-16}; **Figure 2C**). The combined model achieved a sensitivity of 84% and specificity of 68% (Methods). The GRS represents a life-long predisposition to higher urate levels and can be calculated at birth without measurement of

serum urate. As opposed to synovial fluid analysis or CT-based imaging to diagnose gout, the GRS is less invasive and avoids radiation exposure, Thus, the GRS may have utility to identify individuals with a high genetic predisposition for gout, allowing for compensatory lifestyle choices to be made earlier in life to reduce the risk of developing gout.

High genetic correlations of serum urate with cardio-metabolic traits

Serum urate is positively correlated with many cardio-metabolic risk factors and diseases[32]. We assessed genetic correlations between serum urate and 748 complex traits using cross-trait LD score regression (Methods). Serum urate levels were significantly ($p<6.6x10^{-5}=0.05/748$) genetically correlated with 214 complex traits and diseases (**Supplementary Table 12**). The highest positive genetic correlation (r_g) was with gout ($r_g=0.92$, $p=3.3x10^{-70}$), followed by traits representing components of the metabolic syndrome such as HOMA-IR ($r_g=0.49$) and fasting insulin ($r_g=0.45$). Significant positive genetic correlations were also observed for other cardio-metabolic traits or diseases, including waist circumference, obesity, and type 2 diabetes (**Figure 3**). The largest negative correlations were observed with HDL cholesterol-related measurements (r_g up to -0.46), and with estimated glomerular filtration rate (r_g=-0.38 and -0.26 for cystatin C-based and creatinine-based eGFR, respectively), consistent with the known role of the kidneys in urate excretion. Overall, the genetic correlations between serum urate and other complex traits and diseases were consistent with observational associations of serum urate levels with cardio-metabolic traits in epidemiological studies[32].

To examine whether these genetic correlations reflect causal relationships or pleiotropy, we applied a recently-developed latent causal variable model to estimate the genetic causality proportion (GCP) for seven commonly-studied cardio-metabolic traits (Methods). As a positive control, we analyzed gout, confirming a genetically causal effect of urate on gout (GCP=0.79, **Supplementary Table 13**). Conversely, we identified a range of GCP values consistent with mostly or partially genetically causal effects of the seven cardio-metabolic traits on serum urate levels. The highest GCP estimates were observed for adiposity-related traits (e.g. GCP=0.84 for waist circumference, **Supplementary Table 13**), consistent with higher cell numbers resulting in

higher production of purines and consequently urate, as well as with a Mendelian Randomization study that reported a causal effect of adiposity on urate levels.[33] HDL cholesterol levels, on the other hand, showed smaller GCP estimates (GCP<0.5; **Supplementary Table 13**), suggesting the existence of a genetic process with a causal effect on both HDL cholesterol and urate. A potential explanation for such a process are co-regulated metabolic processes in the liver that influence both cholesterol and urate levels. These processes may explain a large fraction of heritability for cholesterol levels and a modest fraction of heritability for urate, a type of asymmetry expected to produce a partially genetically causal relationship consistent with the one observed.

Identification of enriched tissues and pathways

To identify molecular mechanisms and tissues relevant for urate metabolism and handling, and to provide potential clues to the observed genetic correlation with other traits and diseases, we investigated which tissues, cell types and systems may be significantly enriched for the expression of genes mapping into the urate-associated loci. Based on all SNPs with $P<1x10^{-5}$ (Methods), we identified significant enrichment (false discovery rate [FDR] <0.01) for 19 physiological system entries, three tissues, and two cell types (**Supplementary Table 14**). The strongest enrichment was observed for kidney ($P=9.5x10^{-9}$) and urinary tract ($P=9.9x10^{-9}$), both within the urogenital system, consistent with the kidney's prominent role in controlling serum urate concentrations. Additional significant enrichments were observed in the endocrine and digestive system, including liver, the major site of urate production. Interestingly, a novel significant enrichment was also observed in the musculoskeletal system, specifically for synovial membrane, joint capsule, and joints (**Figure 4A**), the sites of gout attacks.

We next tested for cell-type groups with evidence for enriched heritability based on cell-type specific functional genomic elements using stratified LD score regression (Methods). The strongest heritability enrichment was observed for kidney (11.5-fold), followed by liver (5.39-fold; **Supplementary Table 15**). This approach complemented the gene expression-based approach and also supported kidney and liver as the major organs of urate homeostasis.

Lastly, we tested whether any gene sets were enriched for variants associated with urate at $P<10^{-5}$ (Methods). Significant enrichment (FDR <0.01) was observed for 383 reconstituted gene sets (**Supplementary Table 16**). Since many of these contained overlapping groups of genes, we used affinity propagation clustering to identify 57 meta gene sets (Methods, **Supplementary Table 17**), including a prominent group of inter-correlated gene sets related to kidney and liver development, morphology and function (**Figure 4B**). Together, these analyses underscore the prominent role of the kidney and liver in regulating serum urate levels and implicate the kidney as a major target organ for lowering of serum urate levels.

Prioritization of urate loci based on statistical fine-mapping, functional annotation, and gene expression

To prioritize target SNPs and genes for translational research, we established a workflow that combined fine-mapping of urate-associated loci with functional annotation and a systematic evaluation of tissue-specific differential gene expression.

Statistical fine-mapping prioritizes candidate SNPs

To identify independent and potentially causal variants, statistical fine-mapping was performed starting from the 123 genome-wide significant loci identified in the EA-specific meta-analysis, because the workflow included methods that used LD estimates from an ancestry-matched reference panel (Methods)[34]. After combining the 123 loci into 99 larger genomic regions based on LD between index SNPs, stepwise model selection in each region identified 114 independent SNPs ($r^2<0.01$, Methods). Overall, 87 regions contained one independent signal, ten contained two independent SNPs, the *ABCG2* locus contained three and the *SLC2A9* locus four independent SNPs (**Supplementary Table 18**). For each of these 114 independent SNPs, we computed 99% credible sets representing the smallest set of SNPs which collectively account for 99% posterior probability of containing the variant(s) driving the association signal[35]. The 99% credible sets contained a median of 16 SNPs (Q1, Q3: 6, 57), and six of them only a single SNP, mapping in or near *INSR*, *RBM8A*, *MPPED2*, *HNF4A*, *CPT1C*, and *SLC2A9* (**Supplementary Table 18**). Among 28 small credible sets (≤5 SNPs), several mapped in or near genes with an

593 established role in regulating urate levels such as *SLC2A9*, *PDZK1*, *ABCG2*, *SLC22A11*, and
594 *SLC16A9*[20]. These credible sets contain the most supported candidate SNPs based on the serum
595 urate association signals and would greatly reduce the number of candidate functional variants
596 to be tested in experimental follow-up studies.

597 To further refine the credible set SNPs, we annotated them with respect to their
598 functional consequence and regulatory potential (Methods). Missense SNPs with posterior
599 probabilities >50% for driving the association signals or mapping into small credible sets were
600 identified in *ABCG2*, *UNC5CL*, *HNF1A*, *HNF4A*, *CPS1*, and *GCKR* (**Supplementary Table 19, Figure**
601 **5A**). All missense SNPs except the one in *GCKR* had a CADD score >15, supporting the SNP and
602 the gene it resides in as potentially causal. Indeed, functional effects have already been
603 demonstrated experimentally for variants rs2231142 (Gln141Lys, r^2=1 to the index SNP
604 rs74904971) in *ABCG2*, rs742493 (p.Arg432Gly) in *UNC5CL*, and rs1260326 (p.Leu446Pro) in
605 *GCKR* (**Table 1**). Non-exonic variants with posterior probabilities of >90% and mapping into
606 open chromatin in enriched tissues were identified in *RBM8A*, *SLC2A9*, *INSR*, *HNF4A, PDZK1*,
607 *NRG4*, *UNC5CL*, and *AAK1* (Methods, **Supplementary Figure 7**, **Supplementary Table 19**). When
608 complemented by evidence of gene expression co-localization, these SNPs may represent
609 causal regulatory variants and highlight their potential effector genes.

610 *Gene prioritization via gene expression co-localization analyses*

611 To systematically assess differential gene expression, we tested for co-localization of the urate
612 association signals with expression quantitative trait loci (eQTL) in *cis* across three kidney tissue
613 resources and 44 GTEx tissues (Methods). High posterior probability of co-localization (H4≥0.8,
614 Methods) supports a trait-associated variant acting through modulation of gene expression in
615 the tissue where co-localization is identified. The eQTLs from the three kidney tissue resources
616 were based on glomerular and tubulo-interstitial portions of micro-dissected kidney biopsies
617 from 187 CKD patients and healthy kidney tissue sections of 96 additional individuals
618 (Methods). We identified co-localization with the expression of 13 genes in at least one kidney
619 tissue (**Figure 6**), the tissue with the strongest enrichment for urate-associated variants.
620 Whereas co-localization of some genes was only observed in kidney (*SLC17A4*, *BICC1*, *UMOD*,

GALNTL5, *NCOA7*), other genes showed co-localization across several tissues (e.g., *ARL6IP5*). The direction of change in gene expression with higher urate levels could vary for the same gene across tissues. For instance, the allele associated with higher serum urate at the *SLC16A9* locus was associated with higher gene expression in kidney, consistent with a regulatory variant in a transporter mediating the reabsorption of urate. The same allele was associated with lower gene expression in other tissues such as aorta, pointing towards tissue-specific regulatory mechanisms[36]. Details on each of the 13 genes with high posterior probability of a variant underlying the associations with both urate and gene expression in kidney are summarized in **Supplementary Table 20.** Significant co-localizations identified across all 47 tissues (**Supplementary Figure 8**) revealed additional novel insights such as co-localization of the urate association signal with *NFAT5* expression in subcutaneous adipose tissue emphasizing its role in adipogenesis[37], or *PDZK1* expression in colon and ileum, important sites of urate excretion.

Lastly, we investigated whether any trans-ethnic index SNPs or their proxies (r^2>0.8) were reproducibly associated with gene expression in *trans* in whole blood or peripheral blood mononuclear cell in several large eQTL studies (**Supplementary Table 21**, Supplementary Information). We identified inter-chromosomal associations between five index SNPs and 16 transcripts, that were enriched in the term "cardiovascular disease" based on the Human Disease Ontology database (**Supplementary Information, Supplementary Table 22**).

***HNF4A* activates *ABCG2* transcription and HNF4A p.Thr139Ile is a functional variant**

The gene and variant prioritization workflow was validated using the identified candidates *HNF1A* and *HNF4A*. Co-regulation of target genes by these master regulators of transcription in kidney proximal tubule cells and liver could potentially explain observed genetic correlations[38].

We first tested whether *HNF1A* and *HNF4A* have the potential to affect transcription of the *ABCG2* gene, which encodes for a urate transporter of major importance in humans and represented the locus with the highest risk for gout in our screen. *ABCG2* contains several predicted HNF1A and HNF4A binding sites in its promoter region (**Figure 5B**). A luciferase reporter assay in the human embryonic kidney cell line HEK 293 was used to assess

transactivation of the human *ABCG2* promoter by HNF4A and HNF1A proteins (Methods, **Supplementary Figure 9A**). Co-expression of HNF4A significantly increased the *ABCG2* promoter-driven luciferase activity, and the activation was dependent on the transfected *HNF4A* expression vector dose and corresponding levels of HNF4A protein (**Figure 5C, Supplementary Figure 9B**). As expected, no increase of luciferase activity occurred with the pGL4 vector without the *ABCG2* promoter that was used as a negative control (**Supplementary Figure 9D and 9E**). Results for HNF1A indicated that the observed association of this locus with serum urate is unlikely to occur via activation of *ABCG2* in kidney cells (**Figure 5C**), but *HNF1A* has been reported to activate transcription of *PDZK1*, a regulatory protein for several other renal urate transporters[39,40] and also identified in this study.

Next, we tested the functional relevance of the prioritized missense p.Thr139Ile allele in HNF4A (NM_178849.2, isoform 1, Methods). Its location within the hinge/ DNA binding domain (**Figure 5D, Supplementary Figure 9F**) supports potentially altered interactions with targeted promoter regions. The isoleucine to threonine substitution at position 139 significantly increased the transactivation of the *ABCG2* promoter and commensurate luciferase activity as compared to the wild-type threonine (**Figure 5E**), without altering HNF4A protein abundance (**Supplementary Figures 9C**). Thus, HNF4A can activate *ABCG2* transcription in a kidney cell line, and HNFA4 p.Thr139Ile is a functional variant. Increased activation of the urate excretory protein ABCG2 by the allele encoding the isoleucine residue should result in lower serum urate levels, consistent with the observed negative association in our GWAS.

Discussion

This large trans-ethnic GWAS meta-analysis of serum urate levels based on 457,690 individuals represents a four-fold increase in sample size over previous studies[21,22,41] and resulted in the identification of 183 urate-associated loci, 147 of which are novel. A genetic urate risk score led to significant improvements of gout risk prediction among 334,880 independent persons, 3.5% of whom had a risk of gout comparable to a Mendelian disease effect size. Genetic correlation

and causality analyses confirmed the causal effect of urate on gout, and were consistent with transcriptional co-regulation as a source of pleiotropy in the widespread genetic correlations between serum urate and cardio-metabolic traits. Tissue and cell type-specific enrichment analyses supported kidney and liver, the sites of urate excretion and generation, as key target tissues. Comprehensive fine-mapping and co-localization analyses with gene expression across 47 tissues delivered an extensive list of target genes and SNPs for follow-up studies, of which we experimentally confirmed *HNF4A* p.Thr139Ile as a functional allele involved in transcriptional regulation of urate homeostasis.

Major challenges of GWAS are to pinpoint causal genes and variants, and to provide actionable insights into disease-relevant mechanisms and pathways in order to improve disease treatment and prevention. This study developed a comprehensive resource of candidate SNPs, genes, tissues and pathways involved in urate metabolism that will enable a wide range of follow-up studies such as our proof-of-principle validation study. Out of the many novel and biologically plausible findings, we highlight three instances in which co-localization of the serum urate and tissue-specific gene expression signals provides new insights into urate metabolism and the prominent role of the renal proximal tubules. First, co-localization helped to prioritize genes in association peaks that previous GWAS could not resolve. For example, the association signal at chromosome 6p22.2 contains the genes encoding four members of the SLC17 transporter family (*SLC17A1*, *SLC17A2*, SLC17A3, and *SLC17A4*). Systematic testing of co-localization across genes and tissues supported a variant underlying the urate association signal and differential gene expression only for *SLC17A4* in kidney, with higher expression associated with higher serum urate. Previous experimental studies have implicated SLC17A4 as a urate exporter in intestine[42], and our data support its yet unappreciated role in urate transport in the human kidney. Second, co-localization with gene expression provided insights into transport processes of the proximal tubule, the major site of urate reabsorption and excretion. For example, the gene product of the candidate *ARL6IP5* has been shown to modulate activity of the glutamate transporter SLC1A1[43,44], dysfunction of which causes aminoaciduria[45]; and deletion of the candidate *NCOA7* in mice results in distal renal tubular acidosis[46]. Third, it is noteworthy that co-localization of the urate association signal was observed with differential

707 expression of *MUC1*, *BICC1* and *UMOD* in kidney. Rare mutations in all three genes are known
708 to cause monogenic cystic kidney diseases[47-49], pointing towards a shared mechanism with
709 respect to their association with urate.

710 Another noteworthy finding from this well-powered study was the significant genetic
711 correlations with many cardio-metabolic traits, with directions matching expectation from
712 known observational associations[50]. Many of these traits, like lipid levels and urate, are
713 influenced by liver metabolism. We estimated the genetic causality proportion for these traits,
714 showing that their genetic correlations are partly driven by overlapping or co-regulated
715 metabolic pathways in the liver and not only by a fully causal effect of cholesterol or insulin
716 levels on urate. Likewise, significant genetic correlations with kidney-related traits such as eGFR
717 may reflect shared regulation of processes in the kidney, the major site of urate excretion.
718 Evidence for transcriptional co-regulation, beyond the known importance of *HNF1A* and *HNF4A*
719 in liver and kidney, is supported by the identification of additional urate loci such as *MLXIPL,*
720 *TCF7L2* and *KLF10* that share associations with other metabolic and/or kidney function traits.
721 The observed pleiotropic effects of many urate-associated variants could thus be the potential
722 manifestation of co-regulation of processes that occur within and across tissues relevant to the
723 implicated traits, a mechanism likely to be prevailing in metabolic, but also other traits.

724 In the kidney, nuclear HNF4A is exclusively detected in epithelial cells of the proximal
725 tubule[51], where it has been reported to regulate the expression of SLC2A9 isoform 1[52] and
726 PDZK1[53]. Kidney-specific deletion of *HNF4A* in mice phenocopies Fanconi renotubular
727 syndrome[54]. Detailed kidney tissues transcriptomic analyses support HNF4A to drive a proximal
728 tubule signature cluster of 221 co-expressed genes, including many candidate genes for urate
729 metabolism and transport[51]. In addition to *HNF4A*, *HNF4G*, and *HNF1A*, ten genes in this cluster
730 of co-expressed genes also map into urate-associated loci identified here (*A1CF, CUBN, LRP2,*
731 *PDZK1, SERPINF2, SLC2A9, SLC16A9, SLC17A1, SLC22A12 and SLC47A1*). In addition, our study
732 establishes that HNF4A can also trans-activate transcription of *ABCG2* in a kidney cell line, the
733 key urate secretory transporter in both gut and kidney epithelium[55]. The T allele, encoding the
734 isoleucine substitution at *HNF4A* T139I, showed higher trans-activation of *ABCG2* transcription
735 compared to the wild-type allele, which should result in increased urate secretion and is

consistent with the observed association of the T allele with lower serum urate levels. The genetic variant encoding the T139I substitution is located in a region of the HNF4A protein harboring many causative mutations for monogenic maturity onset diabetes of the young (MODY type 1)[56]. Yet, unlike the severe MODY1 missense mutations R127W, D126Y, and R125W,[57] T139I has not been reported to cause MODY1. Instead, it has been reported to increase the risk of type 2 diabetes mellitus, possibly through a liver-specific loss of HNF4A phosphorylation at T139, and to associate with HDL-cholesterol levels[56,58]. These data point to additional complexities when interpreting pleiotropic effects, because there may be several tissue-specific mechanisms by which genetic variants in transcriptional regulators influence metabolic pathways and urate homeostasis.

Despite many strengths of this study, some limitations warrant mention. The numbers of individuals of ancestries other than European or East Asian were still small, highlighting the value of studying more diverse populations. Focusing on SNPs present in the majority of studies emphasizes those that may be of greatest importance globally over population-specific variants. General limitations of the field that are not specific to our study include the facts that statistical fine-mapping approaches based on summary statistics from meta-analyses cannot clearly prioritize functional variants in regions of tight LD, and that they are influenced by the availability of and imputation quality of SNPs in the individual contributing studies. Moreover, only few regulatory maps from important target tissues such as synovial membrane and kidney are available, but we were able to evaluate differential gene expression in three separate kidney datasets. The generation of additional regulatory and expression datasets across disease states, developmental stages and more cell types in the kidney and other metabolically active organs constitutes an important research avenue for the future.

In summary, this large-scale genetic association study of serum urate generated an atlas of candidate SNPs, genes, tissues and pathways involved in urate metabolism and its shared regulation with multiple cardio-metabolic traits that will enable a wide range of follow-up studies.

Online Methods

Overview of GWAS methods

We developed an automated analysis workflow to collect and integrate results from 74 GWAS of serum urate from five ancestry groups participating in the CKDGen Consortium. We used a distributive model for study-specific GWAS with meta-analyses conducted centrally. An analysis plan was circulated to all participating studies accompanied by custom shell and R scripts for phenotype generation (https://github.com/genepi-freiburg/ckdgen-pheno). Study-specific GWAS were conducted after centralized review and approval of the phenotype summary statistics. Study-specific GWAS results were checked using GWAtoolbox[59], including p-value inflation, allele frequency distribution, imputation quality, and completeness of genotypes. Custom scripts were used to compare imputed allele frequencies to those of ancestry-matched reference panels and to visualize variant positions. In addition, quality metrics, including the genomic control factor[60], were compared across studies. The participants of all studies provided written informed consent. Each study had its research protocol approved by the corresponding local ethics committee.

Phenotype definition, genotyping and imputation in participating studies

The primary study outcome was serum urate in mg/dl. The laboratory methods for measuring serum urate in each study are reported in **Supplementary Table 1**. Prevalent gout was analyzed as a secondary outcome to examine whether urate-associated SNPs conferred gout risk. Gout cases were ascertained based on self-report, intake of urate-lowering medications, or International Statistical Classification of Diseases and Related Health Problems (ICD) codes for gout (**Supplementary Table 1**).

Each study performed genotyping separately and applied study-specific quality filters prior to phasing and imputation (**Supplementary Table 2**). In each study, haplotypes were estimated using MACH[61], ShapeIT[62], Eagle[63], or Beagle[64]. Imputation of genotypes was conducted using reference panels from the Haplotype Reference Consortium (HRC) version 1.1[65], 1000 Genomes Project (1000G) phase 3 v5 ALL, or the 1000G phase 1 v3 ALL[66] and

ImputeV2[67], minimac3[68], PBWT[69], the Sanger[70], or the Michigan Imputation Server[68]. The imputed genetic dosages were annotated using NCBI b37 (hg19). Each study provided an imputation quality for each variant: ImputeV2 info score, the MACH/minimac RSQ or the SNPTest info score.

Study-specific association analysis

Each study performed ancestry-specific association analysis of serum urate by generating age- and sex-adjusted residuals of serum urate and regressing the residuals on SNP dosage levels, adjusting for study-specific covariates such as study centers and genetic principal components, assuming an additive genetic model. Gout was analyzed as a binary outcome adjusting for age, sex, genetic principal components, and study-specific covariates. Software used for these regression analyses were EPACTS (*q.emmax* for family based studies and *q.linear* otherwise; https://genome.sph.umich.edu/wiki/EPACTS), SNPTest[71], RegScan[72], RVTEST[73], PLINK 1.90[74], Probabel[75], GWAF[76], GEMMA[77], mach2qtl[78] and R. Family-based studies used methods that accounted for relatedness.

Trans-ethnic, ancestry-specific, and sex-stratified meta-analyses

GWAS results from each study were pre-filtered to retain bi-allelic SNPs with imputation quality score >0.6 and minor allele count (MAC) >10 before inclusion into meta-analysis. Fixed effects inverse-variance weighted meta-analysis was performed using METAL[79] with modifications to output higher precision (six decimal places). Genomic control was applied for each study. The genomic inflation factor λ_{GC}[60] was calculated to assess inflation of the test statistics. For each meta-analysis result (trans-ethnic, ancestry-specific, and sex-specific), we excluded SNPs that were present in <50% of the studies and with a total MAC <400. For ancestry-specific meta-analysis, we additionally excluded SNPs with a heterogeneity I^2-statistic[80] >95%. Genome-wide significance was defined as p-value $<5x10^{-8}$. The LD score regression intercept was calculated to assess the evidence for associations driven by population structure[81]. For downstream characterization, 8,249,849 and 8,217,339 autosomal SNPs were retained in the trans-ethnic and European ancestry meta-analysis, respectively. Ancestry-specific meta-analyses were conducted for European ancestry (EA), African Americans (AA), East Asian (EAS) ancestry, and

South Asian (SA) ancestry using the same methods and variant filters as the trans-ethnic meta-analysis.

Secondary meta-analyses were performed separately in men and women, using the same analytical approaches. To test for significant difference of association between males and females, we used a two-sample t-test $t = \frac{\beta_M - \beta_F}{\sqrt{SE_M{}^2 + SE_F{}^2}}$

where β_M and β_F were beta coefficients in males and females, respectively, and SE_M and SE_F were the standard errors among males and females, respectively.

Initial determination and annotation of genome-wide significant loci

For each meta-analysis result, SNPs were assigned to loci by selecting the SNP with the lowest genome-wide p-value as the index SNP and defining the adjacent ±500kb-region as the corresponding locus. This procedure was repeated until no genome-wide significant SNP remained. An ancestry-specific locus was defined as a genome-wide significant locus in an ancestry-specific meta-analysis of which the index SNP did not map into within the ±500 kb intervals of any genome-wide significant loci in the trans-ethnic meta-analysis. Index SNPs were annotated using its position and the nearest gene based on hg19, RefSeq genes, and dbSNP147 downloaded from ftp://hgdownload.soe.ucsc.edu/mysql/hg19/ on March 23rd, 2017.

Proportion of phenotypic variance explained and estimated heritability

The proportion of phenotypic variance explained by index SNPs was calculated as the sum of the variance explained by each index SNP based on this formula: $\beta^2\left(\frac{2p(1-p)}{var}\right)$, where β is the beta coefficient and p is the MAF of the SNP, and var is the phenotypic variance. For this study, we used the variance of the age- and sex-adjusted residuals of serum urate in EA participants of the ARIC study as the estimate of the phenotypic variance (variance=1.767).

Heritability of age- and sex-adjusted urate was estimated using the R package 'MCMCglmm'[82] in the Cooperative Health Research In South Tyrol (CHRIS) study,[83] a

participating pedigree-based study of EA individuals, comprising 4,373 individuals in 186 pedigrees of up to five generations. We estimated overall heritability, heritability excluding index SNPs in the three major urate loci (*SLC2A9*, *ABCG2*, and *SLC22A12*), and heritability excluding index SNPs in all genome-wide significant loci for both the trans-ethnic and EA-specific meta-analyses. Estimates were obtained by running 1,000,000 MCMC iterations (*burn in* = 500,000) based on previously described settings[84]. The difference between the overall heritability and the heritability excluding the index SNPs represents the heritability explained by the significant loci in the present study.

Trans-ethnic meta-regression

Prior to conducting trans-ethnic meta-regression, we applied the same study-specific SNP filters as those applied to the fixed effects trans-ethnic meta-analysis (imputation quality score >0.6 and MAC >10). An additional filter for MAF >0.0025 was also applied to reduce the influence of very rare SNPs that passed the MAC filter in very large studies. Trans-ethnic meta-regression was conducted using the MR-MEGA software package[85], which models ancestry-associated heterogeneity in the allelic effect as a function of principal components (PCs) generated from a matrix of mean pairwise allele frequency differences between studies. Three principal components generated from a matrix of mean pairwise allele frequency differences between studies were sufficient to separate the self-reported ancestry groups. Due to software requirements, the minimum number of cohorts for each SNP had to be greater than the number of PCs plus two, resulting in the exclusion of SNPs present in five or fewer cohorts. The effect and p-value of each SNP on serum urate was reported, along with additional p-values per-SNP for ancestry-associated heterogeneity ($p_{anc\text{-}het}$) and residual heterogeneity ($p_{res\text{-}het}$). Index SNPs from the fixed effects meta-analysis with $p_{anc\text{-}het} < 2.7 \times 10^{-4}$ (0.05/183) in MR-MEGA were considered to have significant ancestry-associated heterogeneity.

Effect of urate-associated index SNPs on gout and risk prediction for gout

To evaluate the association of the trans-ethnic urate-associated index SNPs with gout, we conducted a trans-ethnic meta-analysis of gout with the same study-specific filtering criteria as for the urate trans-ethnic meta-analysis.

The association between a genetic urate risk score constructed from the 114 independent serum urate-associated SNPs identified among European individuals (see fine-mapping section below) and gout was assessed in a large, independent sample from the UKBB (Projects 19655 and 20272)[86]. We selected 334,880 unrelated individuals (pairwise kinship coefficient<0.0313) of White British ancestry with sex chromosome euploidy and concordance of phenotypic and genotypic sex, including 4,908 with gout identified by self-report at the inclusion visit. Individuals with an ICD10 for gout (M10) in hospital admissions who did not self-report gout were excluded from the analysis. A genetic risk score (GRS) was constructed as the sum of the imputed dosage of the allele associated with higher urate levels ("risk alleles") over all SNPs, multiplied by the genetic effect of the risk allele on serum urate levels. The GRS distribution was divided into ten evenly-spaced categories, and individuals assigned to a category based on their GRS. The category with the lowest GRS did not contain any gout cases and so was combined with its adjacent category. Gout status was regressed on GRS category in a logistic model, including age and sex as covariates, with the category containing the largest number of individuals (genetically predicted mean urate levels 4.74-5.02 mg/dl higher compared to individuals without any urate-increasing alleles) as the reference group.

The performance of the GRS for risk prediction of gout was trained on a set containing 90% of the sample and tested on the remaining 10%. Logistic regression was used to regress gout on the GRS alone (genetic model), age and sex (demographic model) and GRS with age and sex (combined model). Each of these models was then used to predict gout status in the test set. Model performance was assessed by comparing predicted and true gout status using Area Under the Curve (AUC) in a Receiver Operating Characteristic (ROC) curve. An optimum cutoff of the ROC curve to report sensitivity and specificity of a combined GRS-based diagnostic test was determined by the maximum of the Youden's index (sensitivity+specificity-1).

Genetic correlation

To assess the genetic correlation between serum urate and other traits in EA, we conducted cross-trait LD score regression[87] using LD Hub[88] with the EA-specific urate meta-analysis results as input. Genetic correlation estimates with 746 traits were obtained from LD Hub, excluding two previous serum urate GWAS results. For presentation, the 212 significantly correlated traits ($p<6.7x10^{-5}$=0.05/746) were grouped into 9 categories based on the trait names and labels and presented in a circos plot.

To determine whether observed genetic correlations between serum urate and cardio-metabolic traits are likely to represent causal relationships, we used the recently-developed latent causal variable (LCV) method to estimate the genetic causality proportion (GCP) between serum urate and another trait.[89] The GCP describes what proportion of the genetic component of one trait also affects the other trait; a positive GCP value indicates that a large proportion of the genetic component of urate also affects the other trait, and vice versa for a negative GCP value. LCV produces posterior mean and standard deviation estimates of the GCP using mixed fourth moments of the bivariate effect size distribution, based on GWAS summary statistics and LD scores. When using summary statistics of cardio-metabolic traits generated from the UKBB, we assumed non-overlapping populations, and overlapping populations otherwise. We selected six unique continuous cardio-metabolic traits commonly examined in epidemiological studies with high genetic correlation with serum urate ($|r_g|>0.35$). We additionally included gout as a positive control and creatinine-based glomerular filtration rate. EA-specific GWAS summary statistics were used as input to match the ancestry of the LD scores used with the method (https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2).

Functional enrichment

To assess gene-set and tissue enrichment, we used the Data-Driven Expression Prioritized Integration for Complex Traits analysis (DEPICT) version 1 release 194[90]. DEPICT performs gene set enrichment analysis by testing whether genes in 14,461 reconstituted gene sets were enriched for urate-associated SNPs. These reconstituted gene sets were generated based on similarity analysis from gene expression of 77,840 samples, manually curated gene-sets,

molecular pathways from protein-protein interaction screening, and gene sets from mouse gene knock-out studies. Additionally, gene expression levels of from 37,427 human microarray samples are compiled in DEPICT and the tissue and cell types were mapped to 209 MeSH first level terms including physiological systems, tissues and cells. Enrichment of tissues and cell types was conducted by assessing whether urate-associated genes are highly expressed in the tissue and cell types.

All variants with urate association p-values $<1x10^{-5}$ from the trans-ethnic meta-analysis results were used as input. Independent SNPs were identified using PLINK 1.9[74] clump command within 500 kb flanking regions and $r^2>0.1$ in the 1000 Genomes phase1 version 3 data excluding the MHC region (chr6:25–35 Mb). False discovery rates (FDRs) were computed using 500 repetitions, and p-values were computed using 5,000 permutations from 500 null GWAS sets adjusting for gene length.

Affinity propagation clustering (APC)[91], implemented in the R package 'APCluster'[92], was applied to all reconstituted gene sets with FDR-corrected enrichment P-value <0.01 to cluster gene sets containing similar combinations of genes. Similarity between sets was assessed by the z-scores of the top ten genes assigned to each gene set. The algorithm identifies a single reconstituted gene set from each cluster as an exemplar (meta gene set) that best represents the reconstituted gene sets within that cluster. A correlation matrix was calculated from the Z-scores of the top ten genes within the meta gene set. Correlations >0.2 were visualized as edges in a network in Cytoscape (http://cytoscape.org).

Stratified LD score regression for functional enrichment

Stratified LD score regression[87] estimates the SNP heritability of urate contributed by the SNPs linked to histone marks in each cell type. The enrichment of a category is defined as the proportion of SNP heritability in that cell type divided by the proportion of SNPs in the same cell type. Here, we assessed urate heritability enrichment in 10 cell types via stratified LD score regression with the EA-specific urate meta-analysis results as the input to match the ancestry of the LD score estimates. The 10 cell types were collapsed from 220 cell type-specific annotations for four histone marks: H3K4me1, H3K4me3, H3K9ac, and H3K27ac. Analyses were also carried

out using trans-ethnic meta-analysis summary statistics as input and results were similar (data not shown).

Statistical fine-mapping of genome-wide significant loci in European ancestry

Statistical fine-mapping to identify potentially causal variants was performed for the genome-wide significant loci from the EA-specific meta-analysis, because fine-mapping based on summary statistics relies on LD estimates from an ancestry-matched reference panel whose sample size should scale with that of the GWAS (Methods)[34]. We estimated LD from genotypes of 15,000 randomly selected UKBB participants (Application ID 2027, Dataset ID 8974). First, we removed individuals who withdrew consent, those with mismatched sex between self-reported and genetic, with genotype call rates <95%, outliers of variant heterozygosity or along the first two principal components from a principal component analysis seeded with the HapMap phase 3 release 2 populations. We retained only one member of each pair of individuals with pair-wise identity-by-descent statistic ≥0.1875. Altogether, 13,558 individuals with 16,969,363 SNPs were selected as the LD reference panel for fine-mapping.

Statistical fine-mapping was conducted by first combining neighboring loci with correlated index SNPs ($r^2 \geq 0.2$) in the EA meta-analysis into discrete regions. Next, we performed stepwise model selection using GCTA (cojo-slct option) to identify independent index SNPs in each region. For regions with more than one independent index SNPs, we performed conditional analysis (GCTA cojo-cond option) to obtain conditional beta and standard errors. Approximate Bayes factors (ABF) were calculated using the Wakefield's formula[35], as implemented in the R package 'gtx' version 2.0.1 (https://github.com/tobyjohnson/gtx) using the conditional betas and standard errors for regions with multiple independent SNPs and the betas and standard errors of the original EA meta-analysis for regions with a single independent index SNP. The prior standard deviation was calculated as 0.061 based on formula (8) of the original publication of the Wakefield's formula (ref) and the 95% interval of the SNP effect sizes in the EA meta-analysis. The posterior probability for a variant being the driver of the association signal was calculated as the ABF of the variant divided by the sum of the ABF in the region. The 99% credible sets of a region is

derived by summing the posterior probabilities in descending order until the cumulative posterior probability was >99%.

Annotation of the variants in the credible sets

We annotated SNPs in the credible sets for their exonic effect, Combined Annotation Dependent Depletion (CADD) score, and mapping into DNaseI-hypersensitive sites (DHS) from the Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics Consortium projects[93,94]. The exonic effect and CADD score were obtained using SNiPA v3.2 (March 2017)[95]. SNiPA presented the CADD score as PHRED-like transformation of the C score, which was based on CADD release v1.3 downloaded from http://cadd.gs.washington.edu/download. A CADD score of 15 is used to distinguish potential pathogenic variants from background noise in clinical genetics, and represents the median value of all non-synonymous variants in CADD v1.0[96,97].

Co-localization analysis of cis-eQTL and urate-associated loci

Co-localization analysis of urate-associated loci with gene expression was conducted using EA-specific urate meta-analysis results, *cis*-eQTL results from micro-dissected human glomerular and tubulo-interstitial kidney portions from 187 individuals in the NEPTUNE study[98], as well as from 44 tissues in the GTEx Project version 6p release[99]. For each locus, we identified all transcripts and all tissue-transcript pairs with reported eQTLs within ±100 kb of each GWAS index SNP. The region for each co-localization test was defined as the eQTL *cis* window in the underlying studies[98,99]. We used the default parameters and prior definitions set in the 'coloc.fast' function from the R package 'gtx' (https://github.com/tobyjohnson/gtx), which is an adapted implementation of Giambartolomei's colocalization method[24]. Evidence for co-localization was defined as H4 ≥0.8, which represents the posterior probability that the association with serum urate and gene expression is due to the same underlying variant. In addition, co-localization of urate-associated loci was also performed with gene expression quantified using RNA sequencing of the healthy tissue portion of 99 kidney cortex samples from the Cancer Genome Atlas (TCGA)[100]. First, all transcripts that shared eQTL variants with urate index SNPs within ±100 kb were extracted. Then the posterior probability of co-localization was

calculated including eQTLs within the *cis*-window (±1Mb from the transcription start site) for each gene using the R coloc package[24] with default values for the three prior probabilities.

Trans-eQTL annotation

We performed trans-eQTL annotation for the 183 trans-ethnic index SNPs and their proxies (r^2 >0.8 in both the 1000 Genomes phase 3 European and East Asian reference panels) in eQTL studies with >1,000 individuals: the Framingham Heart Study[101], Westra et al.[102], Zeller et al.[103], Fehrmann et al.[104], and the LIFE Heart[105] and LIFE-Adult[106] studies (details in Supplementary Methods). To improve stringency of results, we only report inter-chromosomal trans-eQTLs showing gene expression association p-values $<5x10^{-8}$ in at least two of the above mentioned independent samples. To characterize whether the identified trans-associated eQTL transcripts were enriched in any biological pathways, we conducted gene enrichment analysis using DOSE, an R/Bioconductor package for disease ontology semantic and enrichment analysis[107], as well as packages ReactomePA and clusterProfiler for enrichment analysis using the Human Disease Ontology database, GO, KEGG, and Reactome.[108-112] The background included all 19,327 protein-coding genes from Ensembl/Havana reported in Ensembl release 91.

Experimental study

Promoter Binding Site Predictions

For promoter binding site predictions, we used the JASPAR 2018 database[113,114]. The frequency matrices were downloaded for transcription factor binding sites of both vertebrate and human sequences (HNF1A: MA0046.1 and MA0046.2; HNF4A: MA0114.1 and MA0114.2). These matrices were then used to query the promoter region of *ABCG2* (-1285/+362, or base pairs upstream of the transcription start site / and downstream after transcription start site)[115] by means of the LASAGNA 2.0 transcription factor binding site search tool with default parameters and a p-value cutoff of 0.01[116].

Site-Directed Mutagenesis

HNF1A and HNF4A clones were purchased from GeneCopoeia, (EX-A7792-M02 and EX-Z5283-M02 respectively) and were mutagenized using the QuikChange Lightning Site Directed Mutagenesis kit (Agilent Technologies, #210518) per manufacturer's instructions using PAGE purified primers.

HNF1A-A98V-Forward: CCCTGAGGAGGCGGTCCACCAGAAAGCCG;

HNF1A-A98V-Reverse: CGGCTTTCTGGTGGACCGCCTCCTCAGGG;

HNF4A-T139I-Forward: GACCGGATCAGCATTCGAAGGTCAAGC;

HNF4A-T139I-Reverse: GCTTGACCTTCGAATGCTGATCCGGTC.

Luciferase Assay

HEK293T cells were seeded in white-walled 96-well plates coated with Poly-L-lysine at roughly 12,500 cells per well. Cells were transfected 18 hours later with either the ABCG2 promoter (-1285/+362) upstream of a firefly luciferase in the pGL4.14 vector (a generous gift from Douglas D. Ross, University of Maryland School of Medicine), or the pGL4.14 vector (Promega, #E699A) without promoter construct, as well as GFP expressing vector used as an internal negative control (pEGFP-C1, Clontech)[117] using X-tremeGene^{TM} 9 DNA Transfection Reagent (Roche Diagnostics, #6365787001). Transfection cocktails were prepared per manufacturer's specifications either with or without transcription factor using the following ratio: 0.6 µg promoter construct, 0.2, 0.1, or 0.05 µg transcription factor, and 0.05 µg GFP. When no transcription factor was used, pcDNA3.1 was substituted. Approximately 48 hours after transfection, cells were rinsed with 1x PBS, then lysed using Passive Lysis Buffer (Promega #E194A) for 15 minutes. During this incubation, GFP measurements were taken using a CLARIOstar microplate reader (BMG Labtech). Next, 30 µl of Luciferase Reagent (Promega, E297A&B) were added to each well, and the plate was incubated for an additional 20 minutes at room temperature. Finally, luciferase activity was measured using the CLARIOstar microplate reader taking the average over 6 seconds. To evaluate the significance of transactivation of the ABCG2 promoter we compared cells expressing transcription factors to those transfected with the empty vector (pcDAN3.1) and to evaluate TF dose responses or differences in TF variants all

experimental conditions from one plate were compared using an Ordinary one-way ANOVA, accounting for multiple comparisons with a Tukey's multiple comparison test. Statistical analysis was performed using Prism 7 (GraphPad Software Inc, USA).

Western Blots

Equal volumes of deoxycholate-RIPA buffer were added to wells containing desired lysates following the luciferase assay and plates were then incubated at 4°C overnight. Equal volumes of sample + 5x SDS loading dye + 10% β-merceptoethanol were then loaded into 10% Mini-PROTEAN® TGX Stain-Free™ Precast Gels (Bio-Rad, #4568033) and run per manufacturer's specifications. Gels were then cross-linked for 45 seconds and imaged to reveal total protein load, which was used as the loading control for each lane (representative images of these protein gels are found in S.upplementary figure 9). Gels were then transferred onto nitrocellulose membranes using the Trans-Blot® Turbo™ Transfer System (Bio-Rad), blocked for 2 hours at room temperature in 5% milk in TBS-T, and incubated overnight at 4°C with primary antibody. Membranes were then washed 3 times with TBS-T, incubated at room temperature for 1 hour with Donkey anti-rabbit secondary antibody (Jackson ImmunoResearch, #111-035-144) diluted 1:5000 in 2.5% milk in TBS-T. Membranes were then washed again and developed using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Scientific, #34577) and imaged on the ChemiDoc MP imaging system (Bio-Rad). All primary antibodies were diluted 1:1000 in 2.5% milk in TBS-T. Antibodies used included HNF4α (Cell Signaling Technology, #3113) and HNF1α (Cell Signaling Technology, #89670). Antibodies were validated using lysates of overexpressing HEK293T cells transfected with either HNF construct, demonstrating bands at the appropriate sizes (Supplementary Figure 9).

Table 1: Genes implicated as causal via identification of missense variants with high probability of driving the urate association signal. Genes are included if they contain a missense variant with posterior probability of association of >50% or mapping into a small credible set (≤5 variants).

Gene	SNP	#SNPs in set	SNP PP	consequence	CADD	DHS	Gout p-value (EA)	Brief summary of literature and gene function
ABCG2	rs2231142	4	0.41	p.Gln141Lys (NP_004818.2)	18.2	ENCODE epithelial	1.21E-290	Encodes a xenobiotic and high-capacity urate membrane transporter expressed in kidney, liver and gut. Causal variants have been reported for gout susceptibility (#138900) and the Junior Jr(a-) blood group phenotype (#614490). The locus was first identified in association with serum urate through GWAS (PMID:18834626) and confirmed in many studies since. The common causal variant Q141K has been experimentally confirmed (PMID:19506252) as a partial loss of function.
UNC5CL	rs742493	4	0.95	p.Arg432Gly (NP_775832.2) (within Death domain)	21.0	ENCODE epithelial	2.73E-01	Encodes for the death-domain-containing Unc-5 Family C-Terminal-Like membrane-bound protein. Suggested as a candidate gene for mucosal diseases, with a role in epithelial inflammation and immunity (PMID:22158417). Experiments using human HEK293 cells showed that UNC5CL can transduce pro-inflammatory programs via activation of NF-□B, with the 432Gly variant less potent to do so than the 432Arg one (PMID:22158417).
HNF1A	rs1800574	2	0.92	p.Ala98Val (NP_000536.5)	23.4		1.83E-02	Encodes a transcription factor with strong expression in liver, guts and kidney. Rare mutations cause autosomal-dominant MODY type III (#600496). Locus found in GWAS of T2DM (PMID:22325160) and blood urea nitrogen (PMID:29403010). Together with HNF4-alpha, it was first recognized as master regulator of hepatocyte and islet transcription. Knockout mice show proximal tubular dysfunction (Fanconi syndrome). HNF1A enhanced promoter activity of PDZK1, URAT1, NPT4 and OAT4 in human renal proximal tubule cell-based assays (PMID:28724612), supporting a role in the coordinated expression of components of the urate "transportosome".
HNF4A	rs1800961	1	1.00	p.Thr139Ile (NP_000448.3)	24.7	ENCODE pancreas	7.43E-03	Encodes another nuclear receptor and transcription factor that controls expression of many genes, including *HNF1A* and other overlapping target genes. Rare mutations cause autosomal-dominant MODY type I (#125850) and autosomal-dominant renal Fanconi syndrome 4 (# 616026). Shown to regulate expression of SLC2A9 and other members of the urate "transportosome" in cell-based assays (PMID 25209865, PMID:30124855). The GWAS locus has been reported for multiple cardio-metabolic traits and T2DM (PMID:21874001).
CPS1	rs1047891	84	0.84	p.Thr1412Asn (NP_001116105.1)	22.1		5.66E-02	Encodes mitochondrial carbamoyl phosphate synthetase I, which catalyzes the first committed step of the urea cycle by synthesizing carbamoyl phosphate from ammonia, bicarbonate, and 2 molecules of ATP. Rare mutations cause autosomal-recessive carbamoylphosphate synthetase I deficiency (#237300). In addition to hyperammonemia, this disease features increased synthesis of glutamine, a precursor of purines. Elevated uric acid excretion has been reported in patients with hyperammonemia (PMID:6771064). GWAS locus for eGFR (PMID:26831199), homocysteine (PMID:23824729), urinary glycine concentrations (PMID: 26352407).
GCKR	rs1260326	2	0.67	p.Leu446Pro (NP_001477.2)	0.1	ENCODE kidney	4.09E-41	Encodes a regulatory protein prominently expressed in the liver that inhibits glucokinase. Identified in previous GWAS of urate (PMID:23263486) and multiple other cardio-metabolic traits. The 446L protein was shown to be less activated than 446Pro by physiological concentrations of fructose-6-phosphate, leading to reduced glucokinase inhibitory ability (PMID:19643913).

Abbreviation: PP, posterior probability; DHS, DNase-I hypersensitivity site; CADD, Combined Annotation Dependent Depletion phred score; EA, European ancestry.

Acknowledgements

We thank Daniele Di Domizio (Eurac Research) for IT assistance, and Toby Johnson (GSK) for sharing his code and discussion on credible set fine-mapping and co-localization analysis. This research has been conducted using the UK Biobank Resource under Application Number 20272. Study-specific acknowledgements and funding sources are listed in the **Supplementary Information**.

Disclaimer

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of Health, or the US Department of Health and Human Services.

Data Availability

Genome-wide summary statistics for this study are made publicly available through dbGaP accession number phs000930.v7.p1.

Author contributions

Manuscript writing group: Adrienne Tin, Jonathan Marten, Victoria L. Halperin Kuhns, Yong Li, Matthias Wuttke, Holger Kirsten, Karsten B. Sieber, Chengxiang Qiu, Mathias Gorski, Markus Scholz, Adriana M. Hung, Alexander Teumer, Cristian Pattaro, Owen M. Woodward, Veronique Vitart, Anna Köttgen.

Design of this study: Adrienne Tin, Jonathan Marten, Matthias Wuttke, Mathias Gorski, Christian Fuchsberger, Alexander Teumer, Cristian Pattaro, Owen M. Woodward, Veronique Vitart, Anna Köttgen,

Management of an individual contributing study: Adam S. Butterworth, Adriana M. Hung, Adrienne Tin, Afshin Parsa, Aiko P.J. de Vries, Alan B. Zonderman, Alessandro De Grandi, Andres Metspalu, Andrew A. Hicks, Anke Tönjes, Anna Köttgen, Annette Peters, Antje Körner, Antonietta Robino, Archie Campbell, Belen Ponte, Bernhard K. Krämer, Bettina Jung, Brenda W.J.H. Penninx, Bruce M. Psaty, Caroline Hayward, Carsten A. Böger, Cassandra N. Spracklen, Christian Gieger, Christopher J. O'Donnell, Cornelia M. van Duijn, Cristian Pattaro, Daniela Toniolo, Daniele Cusi, Deborah Mascalzoni, Eric Boerwinkle, Erik Ingelsson, Florian Kronenberg, Gardar Sveinbjornsson, Georg Ehret, Gerard Waeber, Ginevra Biino, Girish N. Nadkarni, Grant W. Montgomery, Harold Snieder, Helena Schmidt, Igor Rudan, J. Michael Gaziano, James F. Wilson, James G. Wilson, Jaspal S. S. Kooner, Jeffrey OConnell, Joachim Thiery, Johanne Tremblay, John B. Whitfield, John C. Chambers, Josef Coresh, Kai-Uwe Eckardt, Karen L. Mohlke, Kari Stefansson, Kevin Ho, Koichi Matsuda, Konstantin Strauch, M. Arfan Ikram, Marcus E. Kleber, Marina Ciullo, Mario Pirastu, Markus Loeffler, Markus Scholz, Martin H. de Borst, Matthias Wuttke, Michael Stumvoll, Michele K. Evans, Michiaki Kubo, Mika Kähönen, Murielle Bochud, Myriam Rheinberger, Nicholas G. Martin, Olivier Devuyst, Olli T. Raitakari, Ozren Polasek, Paolo Gasparini, Peter P. Pramstaller, Peter Vollenweider, Pim van der Harst, Qiong Yang, Rainer Rettig, Reinhold Schmidt, RenÕ©e de Mutsert, Robert J. Carroll, Ron T. Gansevoort, Ruth J.F. Loos, Sarah A. Pendergrass, Sarah H. Wild, Stephan J.L. Bakker, Tamara B. Harris, Terho Lehtimäki, Thomas Perls, Ton J. Rabelink, Uwe Völker, Vilmantas Giedraitis, Vilmundur Gudnason, Weihua Zhang, Wieland Kiess, Winfried März, Wolfgang Koenig, Yong Li, Yuri Milaneschi

Critical review of manuscript: Adam S. Butterworth, Adriana M. Hung, Adrienne Tin, Afshin Parsa, Aiko P.J. de Vries, Alan B. Zonderman, Albert V. Smith, Alexander Teumer, André G. Uitterlinden, Anke Tönjes, Anna Köttgen, Annette Peters, Anselm Hoppmann, Antje Körner, Antonietta Robino, Anubha Mahajan, Audrey Y. Chu, Ayush Giri, Bernhard K. Krämer, Bettina Jung, Boting Ning, Bram Prins, Brenda W.J.H. Penninx, Brigitte Kühnel, Bruce M. Psaty, Caroline Hayward, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Christa Meisinger, Christian Fuchsberger, Christian Gieger, Christopher J. O'Donnell, Cristian Pattaro, Daniel F. Gudbjartsson, Daniela Ruggiero, Deborah Mascalzoni, Dennis O. Mook-Kanamori, Erik Ingelsson, Erwin P. Bottinger, Eulalia Catamo, Florian Kronenberg, Gardar Sveinbjornsson, Ginevra Biino, Giorgia Girotto, Girish N. Nadkarni, Graciela Delgado, Grant W. Montgomery, Harold Snieder, Harry Campbell, , Helgi Jonsson, Hilma Holm, Igor Rudan, Ilja M. Nolte, Ingileif Jonsdottir, Iris M. Heid, James F. Wilson, James G. Wilson, Johanna Jakobsdottir, Johanne Tremblay, John B. Whitfield, Jonathan Marten, Josef Coresh, Kai-Uwe Eckardt, Karen L. Mohlke, Karlhans Endlich, Karsten B. Sieber, Katalin Susztak, Kenneth M. Rice, Kevin Ho, Kjell Nikus, Konstantin Strauch, Laura M. Raffield, Leo-Pekka Lyytikäinen, Leslie A. Lange, Luke J. O'Connor, Man Li, Marcus E. Kleber, Marina Ciullo, Markus Loeffler, Markus Scholz, Martin H. de Borst, Martina La Bianca, Martina Müller-Nurasyid, Mary L. Biggs, Mathias Gorski, Matthias Nauck, Matthias Wuttke, Melanie Waldenberger, Michael H. Preuss, Michele K. Evans, Mika Kähönen, Mike A. Nalls, Myriam Rheinberger, Nicholas G. Martin, Niek Verweij, Nina Hutri-Kähöne, Nisha Bansal, Olivier Devuyst, Olli T. Raitakari, Otis D. Wilson,

Ozren Polasek, Patrick Sulem, Pavel Hamet, Peter K. Joshi, Pim van der Harst, Qiong Yang, Rainer Rettig, Ravchel M. Lewis, Raymond Noordam, RenÕ©e de Mutsert, Ruth J.F. Loos, Sahar Ghasemi, Sala Cinzia Felicita, Salman M. Tajuddin, Sanaz Sedaghat, Sarah A. Pendergrass, Sarah H. Wild, Scott D. Gordon, Shih-Jen Hwang, Shona M. Kerr, Stephan J.L. Bakker, Tamara B. Harris, Teresa Nutile, Terho Lehtimäki, Thibaud S. Boutin, Thomas Meitinger, Todd L. Edwards, Ton J. Rabelink, Unnur Thorsteinsdottir, Uwe Völker, Veronique Vitart, Wei Huang, Winfried März, Wolfgang Koenig, Yong Li, Zhi Yu.

Statistical Methods and Analysis: Albert V. Smith, Alexander Teumer, Anna Köttgen, Anselm Hoppmann, Anubha Mahajan, Audrey Y. Chu, Ayse Demirkan, Ayush Giri, Bettina Jung, Boting Ning, Bram Prins, Brigitte Kühnel, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Chris H. L. Thio, Christian Fuchsberger, Cristian Pattaro, Damia Noce, Daniel F. Gudbjartsson, Edith Hofer, Erika Salvi, Federica Rizzi, Gardar Sveinbjornsson, Ginevra Biino, Graciela Delgado, Holger Kirsten, Ilja M. Nolte, Iris M. Heid, Johanna Jakobsdottir, Johanne Tremblay, Jonathan Marten, Jun Liu, Karsten B. Sieber, Katalin Susztak, Kathleen A. Ryan, Katrin Horn, Kenneth M. Rice, Laura M. Raffield, Leo-Pekka Lyytikäinen, Leslie A. Lange, Man Li, Marco Brumat, Marcus E. Kleber, Maria Pina Concas, Markus Scholz, Martin Gögele, Mary L. Biggs, Masahiro Kanai, Masato Akiyama, Massimiliano Cocca, Mathias Gorski, Matthias Nauck, Matthias Wuttke, Michael H. Preuss, Mike A. Nalls, Myriam Rheinberger, Navya Shilpa Josyula, Nicola Pirastu, Niek Verweij, Nina Mononen, Pashupati P. Mishra, Pavel Hamet, Peter J. van der Most, Peter K. Joshi, Pim van der Harst, Qiong Yang, Raymond Noordam, Rico Rueedi, Robert J. Carroll, Sahar Ghasemi, Salman M. Tajuddin, Sanaz Sedaghat, Sarah A. Pendergrass, Shih-Jen Hwang, Tanguy Corre, Teresa Nutile, Thibaud S. Boutin, Todd L. Edwards, Toomas Haller, Veronique Vitart, Weihua Zhang, Winfried März, Yasaman Saba, Yizhe Xu, Yoichiro Kamatani, Yong Li, Yukinori Okada

Subject Recruitment: Aiko P.J. de Vries, Alan B. Zonderman, Andrej Teren, Andres Metspalu, Anke Tönjes, Anna Köttgen, Archie Campbell, Belen Ponte, Bettina Jung, Blair H. Smith, Brenda W.J.H. Penninx, Carsten A. Böger, Christa Meisinger, Cristian Pattaro, Daniela Ruggiero, Daniele Cusi, David J. Porteous, Erwin P. Bottinger, Florian Kronenberg, Gerard Waeber, Harry Campbell, Helgi Jonsson, Igor Rudan, Isleifur Olafsson, James F. Wilson, James G. Wilson, Jaspal S. S. Kooner, Johan Ärnlöv, Johanne Tremblay, John B. Whitfield, John C. Chambers, Katalin Dittrich, Kjell Nikus, Koichi Matsuda, Marina Ciullo, Michele K. Evans, Michiaki Kubo, Mika Kähönen, Myriam Rheinberger, Nicholas G. Martin, Nina Hutri-Kähöne, Olli T. Raitakari, Ozren Polasek, Patrick Sulem, Peter Vollenweider, Reinhold Schmidt, Renée de Mutsert, Ron T. Gansevoort, Saima Afaq, Sandosh Padmanabhan, Sarah A. Pendergrass, Sarah H. Wild, Simona Vaccargiu, Tanja Poulain, Terho Lehtimäki, Ton J. Rabelink, Vilmundur Gudnason, Wei Huang, Winfried März

Bioinformatics: Albert V. Smith, Anna Köttgen, Anselm Hoppmann, Audrey Y. Chu, Ayush Giri, Benjamin Lehne, Bram Prins, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Christian M. Shaffer, Daniela Baptista, Dennis O. Mook-Kanamori, Edith Hofer, Eric Campana, Erika Salvi, Federica Rizzi, Georg Ehret, Giorgio Pistis, Holger Kirsten, Iris M. Heid, James F. Wilson, Johanna Jakobsdottir, Johanne Tremblay, Jonathan Marten, Karen L. Mohlke, Karsten B. Sieber, Katalin Susztak, Katrin Horn, Leo-Pekka Lyytikäinen, Man Li, Marcus E. Kleber, Maria Pina Concas, Markus Scholz, Massimiliano Cocca, Mathias Gorski, Matthias Wuttke, Michael H. Preuss, Navya Shilpa Josyula, Nicola Pirastu, Pashupati P. Mishra, Pavel Hamet, Peter J. van der Most, Raymond Noordam, Reedik Magi, Rico Rueedi, Robert J. Carroll, Sahar Ghasemi, Sanaz Sedaghat, Sarah A. Pendergrass, Scott D. Gordon, Sven Bergmann, Tanguy Corre, Teresa Nutile, Weihua Zhang, Winfried März, Yasaman Saba, Yizhe Xu, Yong Li, Yuri Milaneschi, Zhi Yu.

Interpretation of Results: Adrienne Tin, Alexander Teumer, André G. Uitterlinden, Anna Köttgen, Ayush Giri, Bettina Jung, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Christian Gieger, Christopher J. O'Donnell, Cristian Pattaro, Helgi Jonsson, Holger Kirsten, Iris M. Heid, Johanne Tremblay, Jonathan Marten, Karen L. Mohlke, Karlhans Endlich, Karsten B. Sieber, Katalin Dittrich, Katalin Susztak, Katrin Horn, Kevin Ho, Luke J. O'Connor, Man Li, Markus Scholz, Mathias Gorski, Matthias Wuttke, Myriam Rheinberger, Niek Verweij, Owen M. Woodward, Pavel Hamet, Pim van der Harst, Sahar Ghasemi, Sanaz Sedaghat, Sarah A. Pendergrass, Shih-Jen Hwang, Veronique Vitart, Victoria L. Halperin Kuhns, Wei Huang, Wolfgang Koenig, Yizhe Xu, Yong Li

Genotyping: Alan B. Zonderman, Alexander Teumer, André G. Uitterlinden, Antje Körner, Archie Campbell, Ayse Demirkan, Blair H. Smith, Brenda W.J.H. Penninx, Caroline Hayward, Carsten A. Böger, Cassandra N. Spracklen, Christian Fuchsberger, Cornelia M. van Duijn, Daniela Baptista, Daniela Ruggiero, Daniela Toniolo, David J. Porteous, Dennis O. Mook-Kanamori, Erik Ingelsson, Erika Salvi, Federica Rizzi, Florian Kronenberg, Georg Ehret, Grant W. Montgomery, Harry Campbell, James F. Wilson, James G. Wilson, Jaspal S. S. Kooner, Johan Ärnlöv, Johanne Tremblay, John C. Chambers, Karen L. Mohlke, Leo-Pekka Lyytikäinen, Leslie A. Lange, Marcus E. Kleber, Melanie Waldenberger, Michael H. Preuss, Michele K. Evans, Michiaki Kubo, Mika Kähönen, Mike A. Nalls, Najaf Amin, Nina Mononen, Olli T. Raitakari, Patrick Sulem, Pavel Hamet, Peter Kovacs, Pim van der Harst, Ralph Burkhardt, Ron T. Gansevoort, Salman M. Tajuddin, Sandosh Padmanabhan, Scott D. Gordon, Simona Vaccargiu, Terho Lehtimäki, Thomas Meitinger, Uwe Völker, Wei Huang, Winfried März, Wolfgang Koenig, Yuri Milaneschi

Functional study: Victoria Halperin Kuhns, Raychel Lewis, Owen M. Woodward.

Competing interests

Dennis O Mook-Kanamori works as a part-time clinical research consultant for Metabolon, Inc. Brenda Penninx has received research funding (unrelated to the work reported here) from Jansen Research and Boehringer Ingelheim. Karsten B. Sieber is full-time employee of GlaxoSmithKline. Gardar Sveinbjornsson and Patrick Sulem are full time employees of deCODE genetics, Amgen Inc. Wolfgang Koenig received modest consultation fees for advisory board meetings from Amgen, DalCor, Kowa, Novartis, Pfizer and Sanofi, and modest personal fees for lectures from Amgen, AstraZeneca, Novartis, Pfizer and Sanofi. Winfried März is employed with Synlab Services GmbH and holds shares of Synlab Holding Deutschland GmbH. Mike A. Nalls is supported by a consulting contract between Data Tecnica International LLC and the National Institute on Aging (NIA), National Institutes of Health (NIH), Bethesda, MD, USA and consults for Illumina Inc., the Michael J. Fox Foundation, and the University of California Healthcare. Kevin Ho disclosed a research and financial relationship with Sanofi-Genzyme. Bruce M. Psaty serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Adam S. Butterworth received grants from MSD, Pfizer, Novartis, Biogen and Bioverativ and personal fees from Novartis. Markus Scholz consults for and received grant support from Merck Serono not related to this project. Anna Köttgen received grant support from Gruenenthal not related to this project. Other authors declare no competing interests.

Figure Legends

Figure 1: Trans-ethnic GWAS meta-analysis identifies 183 loci associated with serum urate

Outer ring: Dot size represents the genetic effect size of the index SNP at each labeled locus on serum urate. Blue band: $-\log_{10}(p)$ for association with serum urate, by chromosomal position (GRCh37 (hg19) reference build). Red line indicates genome-wide significance ($p=5\times10^{-8}$). Blue gene labels indicate novel loci, gray labels loci reported in previous GWAS of serum urate. Green band: $-\log_{10}(p)$ for association with gout, by chromosomal position. Red line indicates genome-wide significance ($p=5\times10^{-8}$). Inner band: Dots represent index SNPs with significant heterogeneity and are color-coded according to its source: green for ancestry-related heterogeneity ($p_{anc\text{-}het} <2.7\times10^{-4}$ [0.05/183]), red for residual heterogeneity ($p_{res\text{-}het} <2.7\times10^{-4}$), and yellow for both ($p_{anc\text{-}het}$ and $p_{res\text{-}het} <2.7\times10^{-4}$). Loci are labeled with the gene closest to the index SNP.

Figure 2: A genetic risk score (GRS) for serum urate improves gout risk prediction. (A) Histogram of the urate GRS among 334,880 European ancestry participants of the UK Biobank. The Y axes show the number of individuals (left) and the prevalence of gout (right), the x-axis shows categories of the urate GRS. The units on the x-axis represent genetically-predicted serum urate levels (mg/dl) compared to individuals without any urate-increasing alleles; **(B)** Y axis displays the age- and sex-adjusted odds ratio of gout by GRS category (X axis), comparing each other category to the most prevalent one; **(C)** Comparison of the receiver operating characteristic (ROC) curves of different prediction models of gout: genetic (GRS only; red), demographic (age + sex; green), and combined (GRS + age + sex; blue). Y-axis: sensitivity, x-axis: specificity. At the optimal cut points determined by the maximum of the Youden's index, the sensitivity of the combined model was 84% and specificity was 68%.

Figure 3: Serum urate shows widespread genetic correlations with cardio-metabolic risk factors and diseases.

The Circos plot shows significant genome-wide genetic correlations between serum urate and 214 complex traits or diseases ($p<6.6x10^{-5}$), with bar height proportional to the genetic correlation coefficient (r_g) estimate for each trait and coloring according to its direction (dark blue, $r_g>0$; light blue, $r_g<0$). Traits and diseases are labeled on the outside of the plot and grouped into nine different categories. Each category is color-coded (inner ring, inset). The greatest genetic correlation was observed with gout ($r_g=0.92$, $p=3.3x10^{-70}$). Genetic correlations with multiple cardio-metabolic risk factors and diseases reflect their known directions from observational studies.

Figure 4: Genes expressed in urate-associated loci are enriched in kidney tissue and pathways. (A) Grouped physiological systems (x-axis) that were tested individually for enrichment of expression of genes in urate-associated loci are shown as a bar plot, with the $-\log_{10}$(*P*-value) on the y-axis. Significantly enriched systems are labeled and highlighted in blue (false discovery rate [FDR] <0.01). **(B)** Correlated (r > 0.2) meta-gene sets that were strongly enriched for genes mapping into urate-associated loci (FDR <0.01). Thickness of the edges represents the magnitude of the correlation coefficient, node size, color and intensity represent the number of clustered gene sets, gene set origin, and enrichment p-value, respectively.

Figure 5: Prioritization of p.Thr139Ile at *HNF4A* and functional study of *HNF4A* regulation of *ABCG2* transcription.

(A) Graph shows credible set size (x-axis) against the posterior probability of association (PPA; y-axis) for each of 1,453 SNPs with PPA >1% in 114 99% credible sets. Triangles mark missense SNPs, with size proportional to their Combined Annotation Dependent Depletion (CADD) score. Blue triangles indicate missense variants mapping into small (≤5 SNPs) credible sets or with high PPA (≥50%). **(B)** Predicted HNF1A or HNF4A binding sites in the promoter region of *ABCG2*, the consensus affinity sequence, and the p-value of likely matches. **(C)** Relative luciferase activity and transactivation of *ABCG2* promoter in cells transfected with variable amount of HNF1A or

HNF4A constructs. ± SEM, n=3 independent experiments, p values calculated with ordinary one-way ANOVA with Tukey's multiple comparison test. **(D)** Position of p.Thr139Ile (T139I) in DNA binding domain / hinge region within HNF4A homodimer structure (PDB 4IQR). **(E)** Relative luciferase activity and transactivation of *ABCG2* promoter in cells transfected with variable amount of constructs (ng's of transfected DNA) of wild-type HNF4A (threonine) or isoleucine at position 139. ± SEM, n=3 independent experiments, p values calculated with ordinary one-way ANOVA with Tukey's multiple comparison test.

Figure 6: Co-localization of urate-association signals with gene expression in *cis* in kidney tissues

Serum urate association signals identified among European ancestry individuals were tested for co-localization with all eQTLs where the eQTL *cis*-window overlapped (±100 kb) the index SNP. Genes with ≥1 positive co-localization (posterior probability of one common causal variant, H4, ≥0.80) in a kidney tissue are illustrated with the respective index SNP and transcript (y-axis). Co-localizations across all tissues (x-axis) are illustrated as dots, where the size of the dots indicates the posterior probability of the co-localization. Negative co-localizations (posterior probability of H4 <0.80) are marked in gray, while the positive co-localizations are color-coded based on the predicted change in expression relative to the allele associated with higher serum urate.

Table 1: Genes implicated as causal via identification of missense variants with high probability of driving the urate association signal. Genes are included if they contain a missense variant with posterior probability of association of >50% or mapping into a small credible set (≤5 variants).

References

1. Kuo, C.F., Grainge, M.J., Zhang, W. & Doherty, M. Global epidemiology of gout: prevalence, incidence and risk factors. *Nat Rev Rheumatol* **11**, 649-62 (2015).
2. Li, X. *et al.* Serum uric acid levels and multiple health outcomes: umbrella review of evidence from observational studies, randomised controlled trials, and Mendelian randomisation studies. *BMJ* **357**, j2376 (2017).
3. Jinno, S., Hasegawa, K., Neogi, T., Goto, T. & Dubreuil, M. Trends in Emergency Department Visits and Charges for Gout in the United States between 2006 and 2012. *J Rheumatol* **43**, 1589-92 (2016).
4. Kuo, C.F., Grainge, M.J., Mallen, C., Zhang, W. & Doherty, M. Rising burden of gout in the UK but continuing suboptimal management: a nationwide population study. *Ann Rheum Dis* **74**, 661-7 (2015).
5. Mikuls, T.R., Farrar, J.T., Bilker, W.B., Fernandes, S. & Saag, K.G. Suboptimal physician adherence to quality indicators for the management of gout and asymptomatic hyperuricaemia: results from the UK General Practice Research Database (GPRD). *Rheumatology (Oxford)* **44**, 1038-42 (2005).
6. Yang, Q. *et al.* Genome-wide search for genes affecting serum uric acid levels: the Framingham Heart Study. *Metabolism* **54**, 1435-41 (2005).
7. Vitart, V. *et al.* SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat Genet* **40**, 437-42 (2008).
8. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet* **2**, e132 (2006).
9. Wang, W. *et al.* Heritability and Genome-Wide Association Analyses of Serum Uric Acid in Middle and Old-Aged Chinese Twins. *Front Endocrinol (Lausanne)* **9**, 75 (2018).
10. MacCluer, J.W. *et al.* Heritability of measures of kidney disease among Zuni Indians: the Zuni Kidney Project. *Am J Kidney Dis* **56**, 289-302 (2010).
11. Rule, A.D. *et al.* Genome-wide linkage analysis for uric acid in families enriched for hypertension. *Nephrol Dial Transplant* **24**, 2414-20 (2009).
12. Enomoto, A. *et al.* Molecular identification of a renal urate anion exchanger that regulates blood urate levels. *Nature* **417**, 447-52 (2002).
13. Li, S. *et al.* The GLUT9 gene is associated with serum uric acid levels in Sardinia and Chianti cohorts. *PLoS Genet* **3**, e194 (2007).
14. Doring, A. *et al.* SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat Genet* **40**, 430-6 (2008).
15. Dehghan, A. *et al.* Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *Lancet* **372**, 1953-61 (2008).
16. Kolz, M. *et al.* Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet* **5**, e1000504 (2009).
17. Yang, Q. *et al.* Multiple genetic loci influence serum urate levels and their relationship with gout and cardiovascular disease risk factors. *Circ Cardiovasc Genet* **3**, 523-30 (2010).
18. Tin, A. *et al.* Genome-wide association study for serum urate concentrations and gout among African Americans identifies genomic risk loci and a novel URAT1 loss-of-function allele. *Hum Mol Genet* **20**, 4056-68 (2011).
19. Woodward, O.M. *et al.* Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc Natl Acad Sci U S A* **106**, 10338-42 (2009).
20. Major, T.J., Dalbeth, N., Stahl, E.A. & Merriman, T.R. An update on the genetics of hyperuricaemia and gout. *Nat Rev Rheumatol* **14**, 341-353 (2018).

21. Kottgen, A. *et al.* Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* **45**, 145-54 (2013).
22. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* **50**, 390-400 (2018).
23. Schaid, D.J., Chen, W. & Larson, N.B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491-504 (2018).
24. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
25. Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet* **42**, 210-5 (2010).
26. Okada, Y. *et al.* Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet* **44**, 904-9 (2012).
27. Merriman, T.R. Population heterogeneity in the genetic control of serum urate. *Semin Nephrol* **31**, 420-5 (2011).
28. Roddy, E. & Choi, H.K. Epidemiology of gout. *Rheum Dis Clin North Am* **40**, 155-75 (2014).
29. Perez-Ruiz, F., Sundy, J.S., Miner, J.N., Cravets, M. & Storgard, C. Lesinurad in combination with allopurinol: results of a phase 2, randomised, double-blind study in patients with gout with an inadequate response to allopurinol. *Ann Rheum Dis* **75**, 1074-80 (2016).
30. Sautin, Y.Y. & Johnson, R.J. Uric acid: the oxidant-antioxidant paradox. *Nucleosides Nucleotides Nucleic Acids* **27**, 608-19 (2008).
31. Long, W. *et al.* Identification of Key Residues for Urate Specific Transport in Human Glucose Transporter 9 (hSLC2A9). *Sci Rep* **7**, 41167 (2017).
32. Feig, D.I., Kang, D.H. & Johnson, R.J. Uric acid and cardiovascular risk. *N Engl J Med* **359**, 1811-21 (2008).
33. Lyngdoh, T. *et al.* Serum uric acid and adiposity: deciphering causality using a bidirectional Mendelian randomization approach. *PLoS One* **7**, e39321 (2012).
34. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am J Hum Genet* **101**, 539-551 (2017).
35. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* **81**, 208-27 (2007).
36. Pao, S.S., Paulsen, I.T. & Saier, M.H., Jr. Major facilitator superfamily. *Microbiol Mol Biol Rev* **62**, 1-34 (1998).
37. Asano, T. *et al.* The role of N-glycosylation of GLUT1 for glucose transport activity. *J Biol Chem* **266**, 24632-6 (1991).
38. Boyle, E.A., Li, Y.I. & Pritchard, J.K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186 (2017).
39. Prestin, K. *et al.* Regulation of PDZ domain-containing 1 (PDZK1) expression by hepatocyte nuclear factor-1alpha (HNF1alpha) in human kidney. *Am J Physiol Renal Physiol* **313**, F973-F983 (2017).
40. Maher, J.M. *et al.* Alterations in transporter expression in liver, kidney, and duodenum after targeted disruption of the transcription factor HNF1alpha. *Biochem Pharmacol* **72**, 512-22 (2006).
41. Sulem, P. *et al.* Identification of low-frequency variants associated with gout and serum uric acid levels. *Nat Genet* **43**, 1127-30 (2011).
42. Togawa, N., Miyaji, T., Izawa, S., Omote, H. & Moriyama, Y. A Na+-phosphate cotransporter homologue (SLC17A4 protein) is an intestinal organic anion exporter. *Am J Physiol Cell Physiol* **302**, C1652-60 (2012).

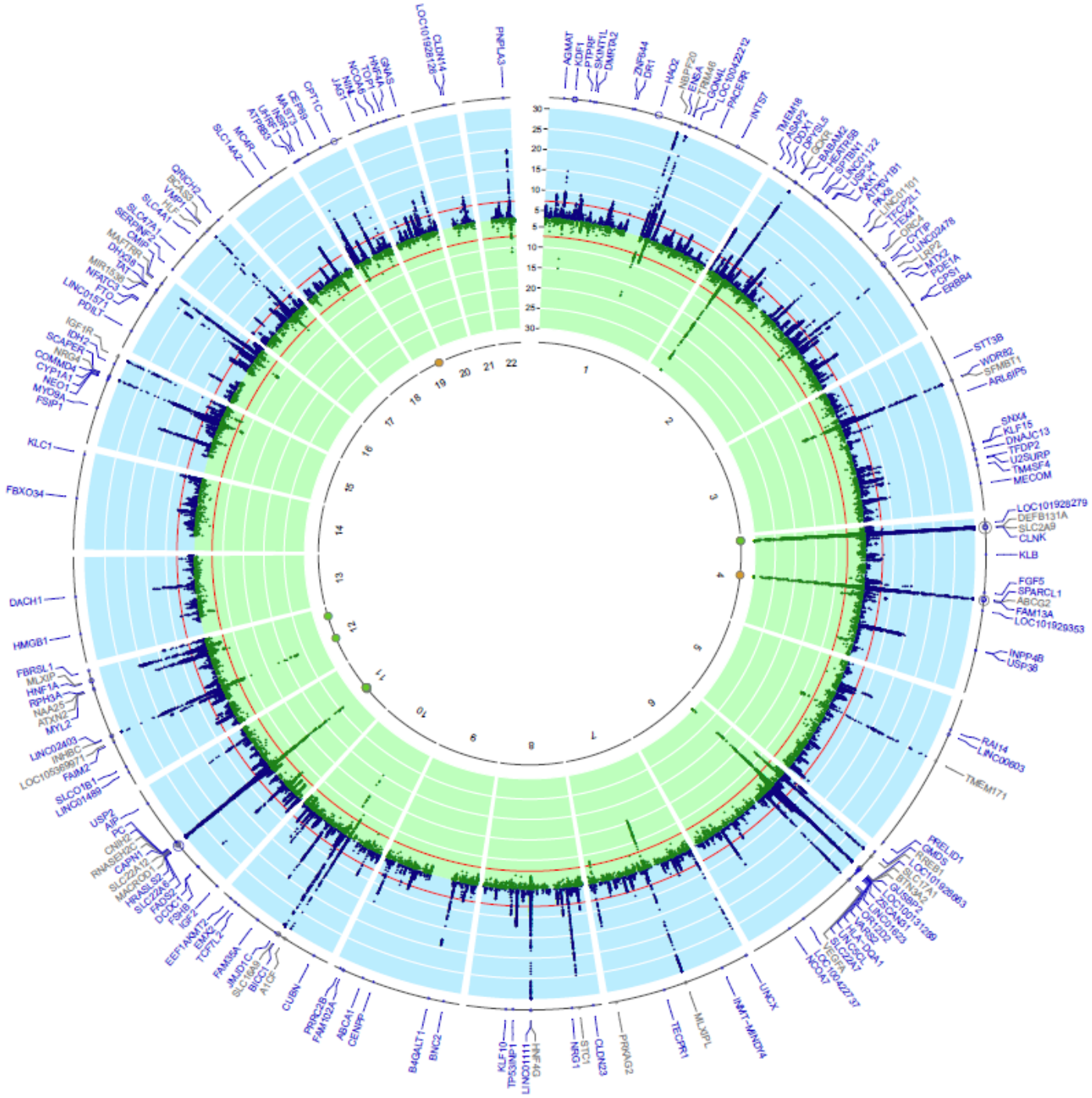
43. Lin, C.I. *et al.* Modulation of the neuronal glutamate transporter EAAC1 by the interacting protein GTRAP3-18. *Nature* **410**, 84-8 (2001).
44. Butchbach, M.E., Lai, L. & Lin, C.L. Molecular cloning, gene structure, expression profile and functional characterization of the mouse glutamate transporter (EAAT3) interacting protein GTRAP3-18. *Gene* **292**, 81-90 (2002).
45. Bailey, C.G. *et al.* Loss-of-function mutations in the glutamate transporter SLC1A1 cause human dicarboxylic aminoaciduria. *J Clin Invest* **121**, 446-53 (2011).
46. Merkulova, M. *et al.* Targeted deletion of the Ncoa7 gene results in incomplete distal renal tubular acidosis in mice. *Am J Physiol Renal Physiol* **315**, F173-F185 (2018).
47. Kirby, A. *et al.* Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. *Nat Genet* **45**, 299-303 (2013).
48. Kraus, M.R. *et al.* Two mutations in human BICC1 resulting in Wnt pathway hyperactivity associated with cystic renal dysplasia. *Hum Mutat* **33**, 86-90 (2012).
49. Hart, T.C. *et al.* Mutations of the UMOD gene are responsible for medullary cystic kidney disease 2 and familial juvenile hyperuricaemic nephropathy. *J Med Genet* **39**, 882-92 (2002).
50. Sodini, S.M., Kemper, K.E., Wray, N.R. & Trzaskowski, M. Comparison of Genotypic and Phenotypic Correlations: Cheverud's Conjecture in Humans. *Genetics* **209**, 941-948 (2018).
51. Lindgren, D. *et al.* Cell-Type-Specific Gene Programs of the Normal Human Nephron Define Kidney Cancer Subtypes. *Cell Rep* **20**, 1476-1489 (2017).
52. Prestin, K. *et al.* Transcriptional regulation of urate transportosome member SLC2A9 by nuclear receptor HNF4alpha. *Am J Physiol Renal Physiol* **307**, F1041-51 (2014).
53. Ketharnathan, S. *et al.* A non-coding genetic variant maximally associated with serum urate levels is functionally linked to HNF4A-dependent PDZK1 expression. *Hum Mol Genet* **27**, 3964-3973 (2018).
54. Marable, S.S., Chung, E., Adam, M., Potter, S.S. & Park, J.S. Hnf4a deletion in the mouse kidney phenocopies Fanconi renotubular syndrome. *JCI Insight* **3**(2018).
55. Matsuo, H. *et al.* ABCG2 dysfunction causes hyperuricemia due to both renal urate underexcretion and renal urate overload. *Sci Rep* **4**, 3755 (2014).
56. Daigo, K. *et al.* Proteomic analysis of native hepatocyte nuclear factor-4alpha (HNF4alpha) isoforms, phosphorylation status, and interactive cofactors. *J Biol Chem* **286**, 674-86 (2011).
57. Chandra, V. *et al.* Multidomain integration in the structure of the HNF-4alpha nuclear receptor complex. *Nature* **495**, 394-8 (2013).
58. Zhu, Q. *et al.* T130I mutation in HNF-4alpha gene is a loss-of-function mutation in hepatocytes and is associated with late-onset Type 2 diabetes mellitus in Japanese subjects. *Diabetologia* **46**, 567-73 (2003).
59. Fuchsberger, C., Taliun, D., Pramstaller, P.P. & Pattaro, C. GWAtoolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics* **28**, 444-5 (2012).
60. Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* **60**, 155-66 (2001).
61. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).
62. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-81 (2011).
63. Loh, P.R., Palamara, P.F. & Price, A.L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat Genet* **48**, 811-6 (2016).

64. Browning, B.L. & Browning, S.R. A fast, powerful method for detecting identity by descent. *Am J Hum Genet* **88**, 173-82 (2011).
65. the Haplotype Reference, C. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* **48**, 1279 (2016).
66. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
67. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
68. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
69. Durbin, R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). *Bioinformatics* **30**, 1266-72 (2014).
70. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
71. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* **11**, 499-511 (2010).
72. Haller, T., Kals, M., Esko, T., Magi, R. & Fischer, K. RegScan: a GWAS tool for quick estimation of allele effects on continuous traits and their combinations. *Brief Bioinform* **16**, 39-44 (2015).
73. Zhan, X., Hu, Y., Li, B., Abecasis, G.R. & Liu, D.J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423-6 (2016).
74. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
75. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
76. Chen, M.H. & Yang, Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* **26**, 580-1 (2010).
77. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-4 (2012).
78. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**, 816-34 (2010).
79. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
80. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539-58 (2002).
81. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
82. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the MCMC glmm R Package. *J Stat Softw* **33**, 1-22 (2010).
83. Pattaro, C. *et al.* The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. *J Transl Med* **13**, 348 (2015).
84. Hadfield, J.D. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software; Vol 1, Issue 2 (2010)* (2010).
85. Magi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* **26**, 3639-3650 (2017).

86. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
87. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
88. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).
89. O'Connor, L.J. & Price, A.L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* **50**, 1728-1734 (2018).
90. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
91. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972-6 (2007).
92. Bodenhofer, U., Kothmeier, A. & Hochreiter, S. APCluster: an R package for affinity propagation clustering. *Bioinformatics* **27**, 2463-4 (2011).
93. Sheffield, N.C. *et al.* Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res* **23**, 777-88 (2013).
94. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
95. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmuller, G. SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334-6 (2015).
96. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* **24**, 2125-37 (2015).
97. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
98. Gillies, C.E. *et al.* An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet* **103**, 232-244 (2018).
99. GTEx. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
100. Ko, Y.A. *et al.* Genetic-Variation-Driven Gene-Expression Changes Highlight Genes with Important Functions for Kidney Disease. *Am J Hum Genet* **100**, 940-953 (2017).
101. Joehanes, R. *et al.* Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* **18**, 16 (2017).
102. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-1243 (2013).
103. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).
104. Fehrmann, R.S. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet* **7**, e1002197 (2011).
105. Ichida, K. *et al.* Decreased extra-renal urate excretion is a common cause of hyperuricemia. *Nat Commun* **3**, 764 (2012).
106. Qu, L.H., Jiang, H. & Chen, J.H. Effect of uric acid-lowering therapy on blood pressure: systematic review and meta-analysis. *Ann Med* **49**, 142-156 (2017).
107. Yu, G., Wang, L.G., Yan, G.R. & He, Q.Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608-9 (2015).
108. Schriml, L.M. *et al.* Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res* **47**, D955-D962 (2019).

109. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res* **45**, D331-D338 (2017).
110. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).
111. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-D361 (2017).
112. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res* **46**, D649-D655 (2018).
113. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* **46**, D1284 (2018).
114. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**, D91-4 (2004).
115. Xie, Y. *et al.* Functional cyclic AMP response element in the breast cancer resistance protein (BCRP/ABCG2) promoter modulates epidermal growth factor receptor pathway- or androgen withdrawal-mediated BCRP/ABCG2 transcription in human cancer cells. *Biochim Biophys Acta* **1849**, 317-27 (2015).
116. Lee, C. & Huang, C.H. LASAGNA-Search 2.0: integrated transcription factor binding site search and visualization in a browser. *Bioinformatics* **30**, 1923-5 (2014).
117. Vesuna, F., Winnard, P., Jr. & Raman, V. Enhanced green fluorescent protein as an alternative control reporter to Renilla luciferase. *Anal Biochem* **342**, 345-7 (2005).

Figure 1


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
30
25
20
15
10
5
5
10
15
20
25
30
AGMAT
KDR
PTPRF
SKINTL
DMRTA2
ZNF844
DR1
HAO2
NBPF20
ENSA
TRM46
GON4L
LOC100422212
PACERR
INTS7
TMEM18
ASAP2
DDX1
DPYSL5
GCKR
BABAM2
HEATR5B
SPTBN1
LINC01122
USP34
AAK1
ATP6V1B1
PAK6
LINC01101
TFCP2L1
TEX41
ORC4
CYP1P
LINC02476
LRP2
MTX2
PDE1A
CPS1
ERBB4
STT3B
WDR82
SFMBT1
ARL6IP5
SNX4
KLF15
DNAJC13
TPDP2
U2SURP
TMEM4F4
MECOM
LOC101928279
DEFB131A
SLC2A9
CLNK
KLB
FGF5
SPARCL1
ABCG2
FAM13A
LOC101929353
INPP4B
USP38
RAI14
LINC00603
TMEM171
FREM1
GMDS
TREB1
LOC101928083
SLC17A1
BTN3A2
GUSBP3
LOC100131289
LINC01623
LRRC4C
LOC105369737
NCOA7
VEGFA
SLC22A7
LINC02572
ORR2D2
HLA-DQA1
UNC5CL
UNC5D
LOC101928083
UNCX
IMMT-MND4
MIJ3PL
TECPR1
PRKAG2
OLDN23
STC1
NRG1
KLF10
TP53INP1
LINC01111
HNF4G
B4GALT1
BNC2
CUBN
PRRC2B
FAM102A
ABCA1
CENPP
FAM83A
JMJD1C
SLC16A6
BICD1
ATG7
EEF1AKNMT2
EMX2
TOPTL2
DCDC1
FSHB
IGF2
FADS2
SLC22A18
HRASLS2
MACROD1
SLC22A12
CAPN1
RNASEH2C
CNIH2
PC
ASB
USP2
SLCO1B1
LINC01489
FAIM2
LOC105369971
INHBC
LINC02403
MYL2
ATXN2
NAA25
RPH3A
HNF1A
MLXIP
FBRSL1
HMGB1
DACH1
FBXO34
KLC1
FSIP1
MYO9A
NEO1
CYP1A1
COMMD4
NRG4
SCAPER
IDH3
IGF1R
PDILT
LINC01571
FTO
NFATC3
MIR1536
DHX38
AT1
MAFTRR
CMIP
SERPINF1
SLC47A1
SLC4A1
HUL
VAMP1
BKAP3
ORICH3
ATP8B3
LHFPL3
MAGI3
OPRM1
CPT1C
JAG1
NR4A1
NCOA6
TOPT
HNF4A
GNAS
CLDN14
LOC101928126
PNPLA3
SLC14A2
MCAR

Track key

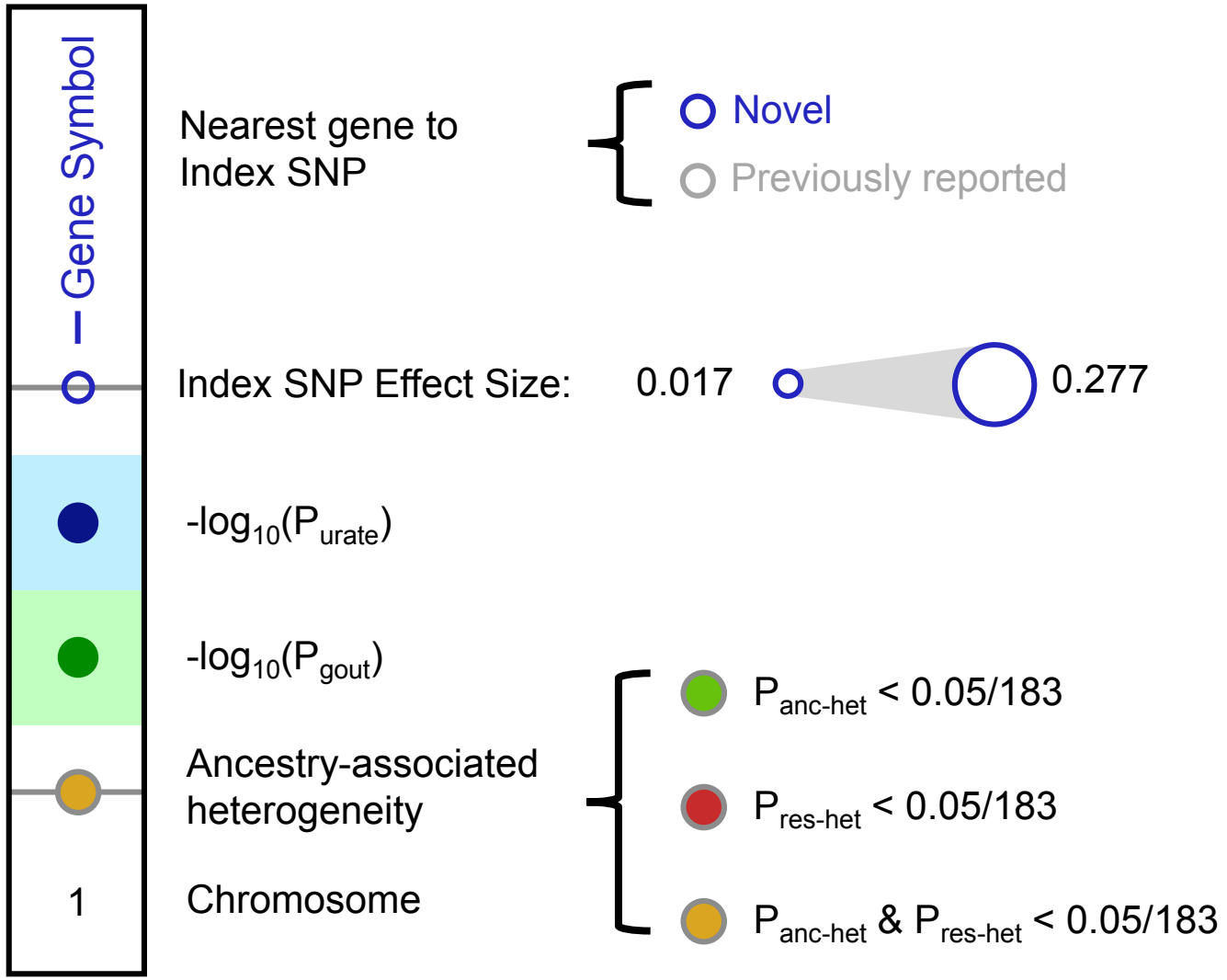
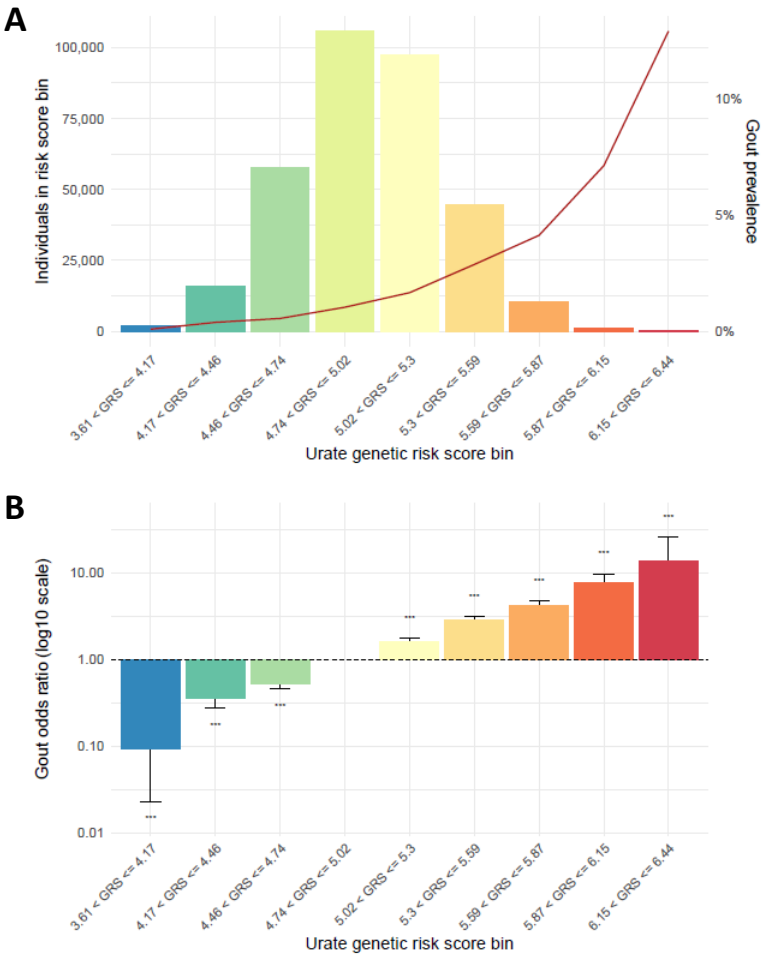

Gene Symbol
Nearest gene to Index SNP
Novel
Previously reported
Index SNP Effect Size:
0.017
0.277
-log10(Purate)
-log10(Pgout)
Ancestry-associated heterogeneity
Panc-het < 0.05/183
Pres-het < 0.05/183
Panc-het & Pres-het < 0.05/183
1
Chromosome

Figure 2


A
Individuals in risk score bin
100,000
75,000
50,000
25,000
0
Gout prevalence
10%
5%
0%
3.61 < GRS <= 4.17
4.17 < GRS <= 4.46
4.46 < GRS <= 4.74
4.74 < GRS <= 5.02
5.02 < GRS <= 5.3
5.3 < GRS <= 5.59
5.59 < GRS <= 5.87
5.87 < GRS <= 6.15
6.15 < GRS <= 6.44
Urate genetic risk score bin

B
Gout odds ratio (log10 scale)
10.00
1.00
0.10
0.01
***
3.61 < GRS <= 4.17
4.17 < GRS <= 4.46
4.46 < GRS <= 4.74
4.74 < GRS <= 5.02
5.02 < GRS <= 5.3
5.3 < GRS <= 5.59
5.59 < GRS <= 5.87
5.87 < GRS <= 6.15
6.15 < GRS <= 6.44
Urate genetic risk score bin

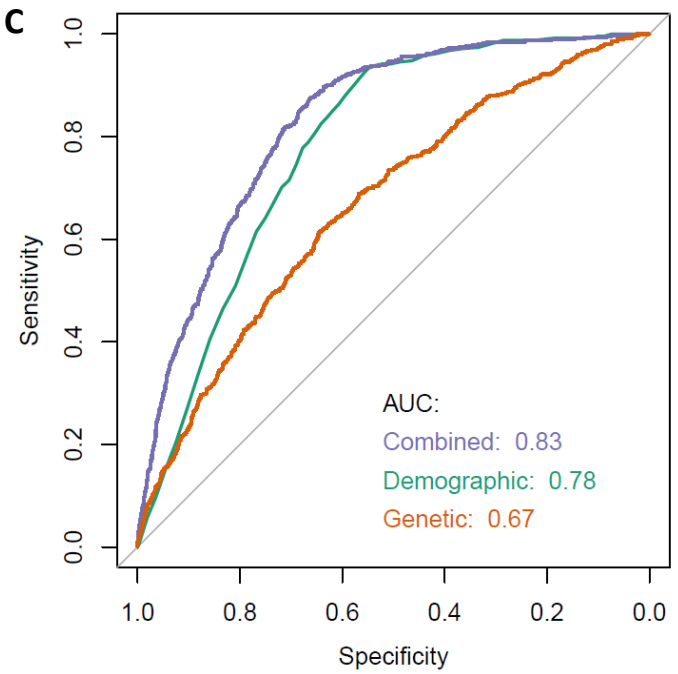
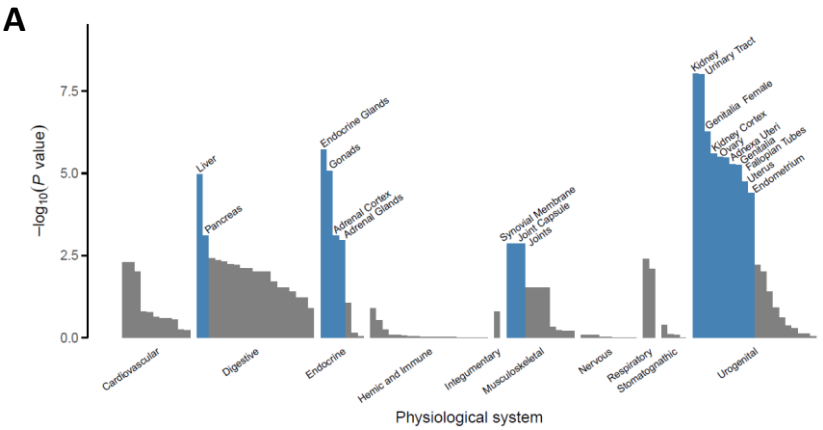
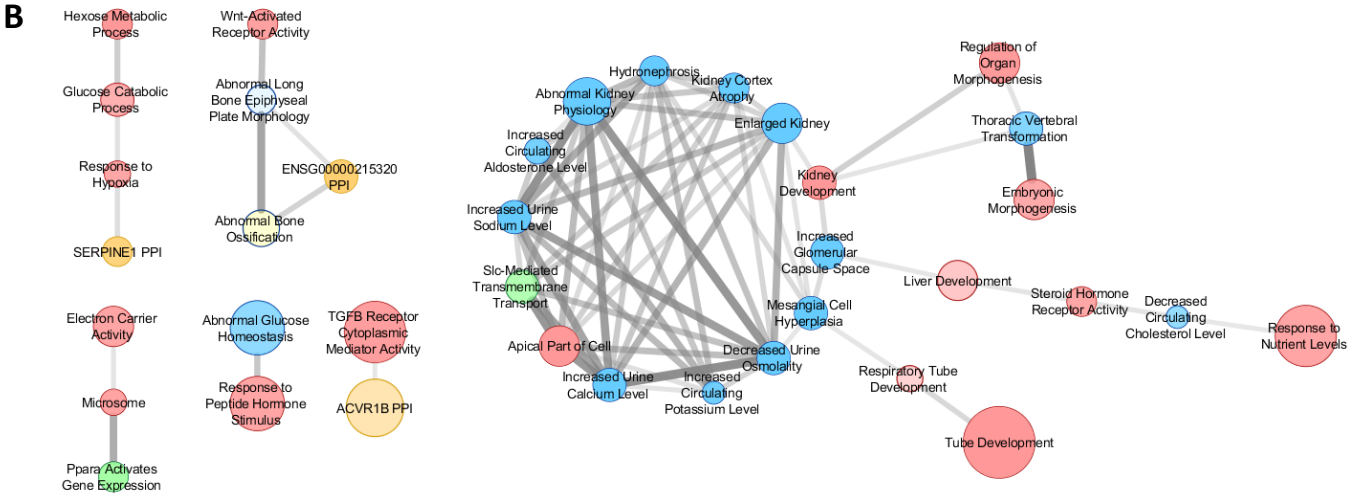

C
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
Specificity
AUC:
Combined: 0.83
Demographic: 0.78
Genetic: 0.67

Figure 3


0.75
0.5
0.25
0
Gout
Alcohol and smoking
Amino acid, protein and lipid
Anthropometric
Cardiovascular and kidney
Education and behavior
Lung function, physical activity
Medication
Sleeping and others
rg > 0
rg < 0

Figure 4





Key:

Gene set origin:
- Gene Ontology
- Mouse Phenotype
- PPI network
- REACTOME pathway

DEPICT P-value: colour intensity

1.68×10^{-10} — 4.5×10^{-4}

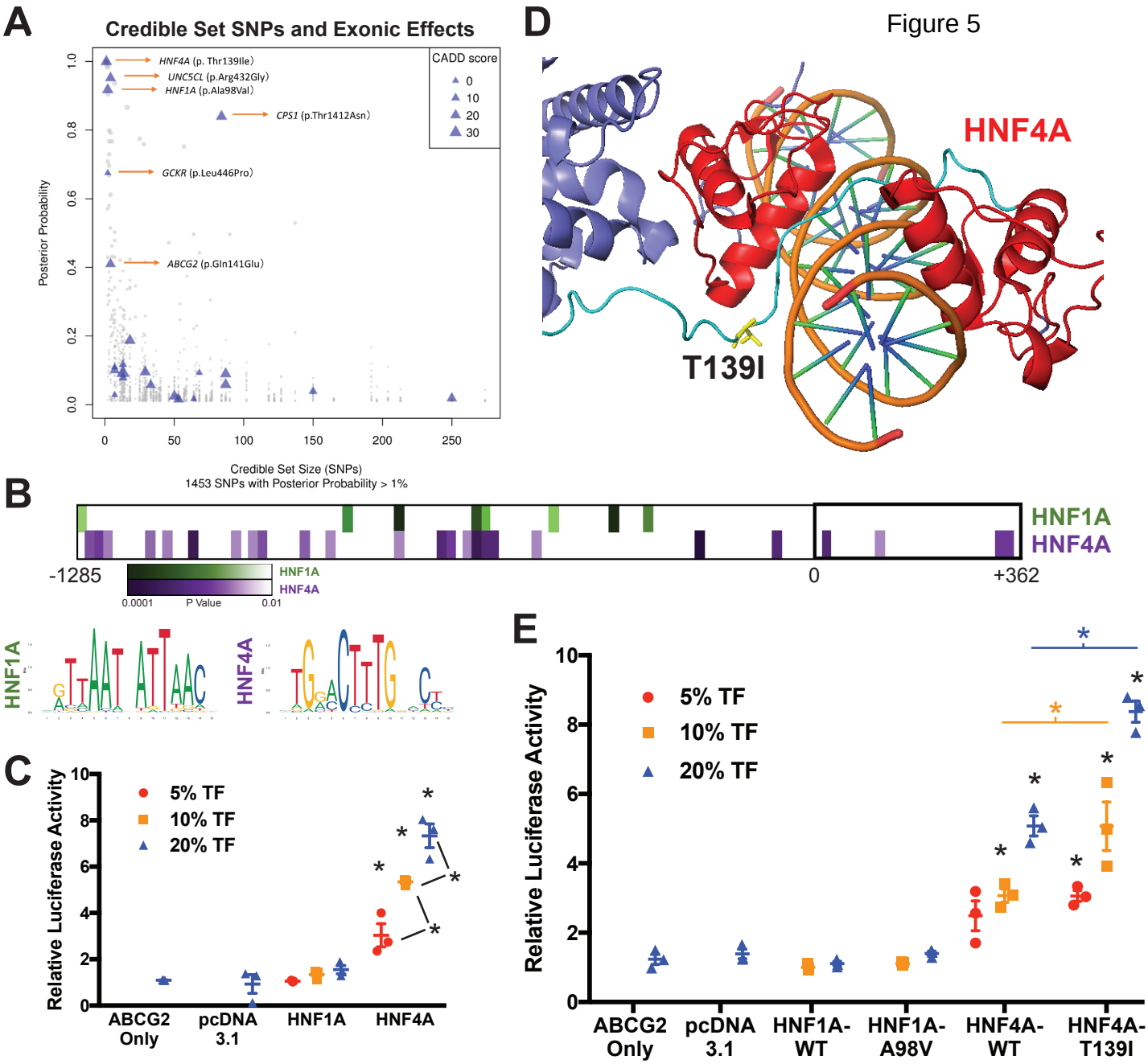
Cluster size: Node size

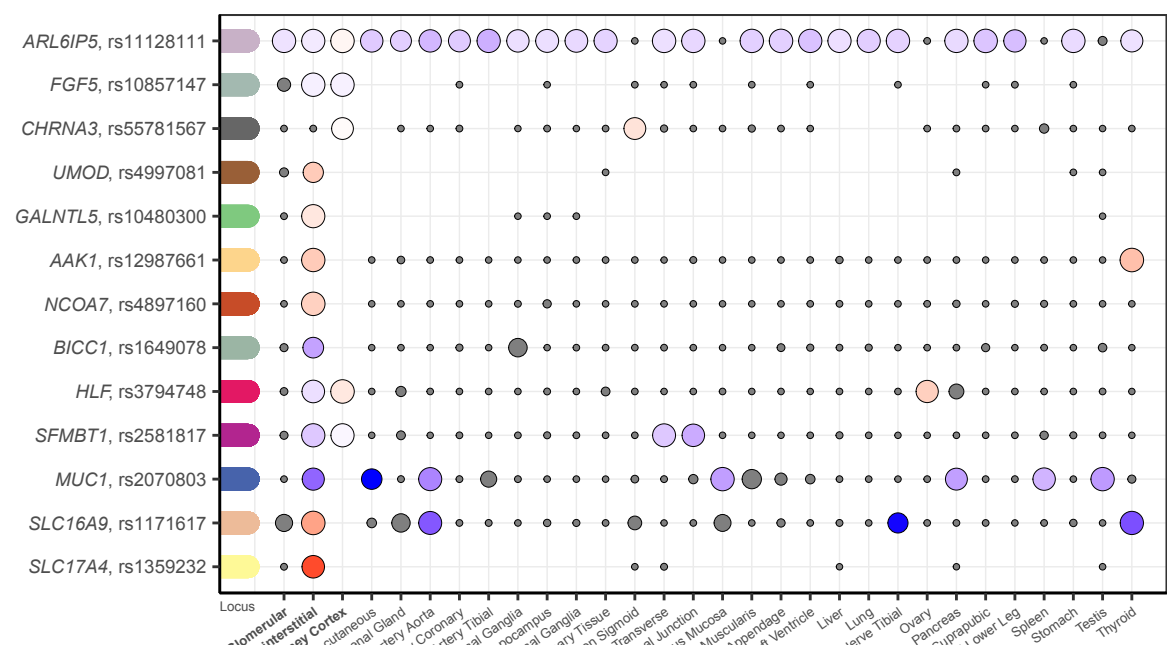
1 — 28

Correlation coefficient: edge

0.2 — 0.89

Figure 5


A
Credible Set SNPs and Exonic Effects
HNF4A (p. Thr139Ile)
UNC5CL (p.Arg432Gly)
HNF1A (p.Ala98Val)
CPS1 (p.Thr1412Asn)
GCKR (p.Leu446Pro)
ABCG2 (p.Gln141Glu)
CADD score
0
10
20
30
Posterior Probability
Credible Set Size (SNPs)
1453 SNPs with Posterior Probability > 1%
D
HNF4A
T139I
B
HNF1A
HNF4A
-1285
0
+362
0.0001
P Value
0.01
C
Relative Luciferase Activity
5% TF
10% TF
20% TF
ABCG2 Only
pcDNA 3.1
HNF1A
HNF4A
E
Relative Luciferase Activity
ABCG2 Only
pcDNA 3.1
HNF1A-WT
HNF1A-A98V
HNF4A-WT
HNF4A-T139I



Posterior probability of colocalization

- [0.0, 0.2)
- [0.2, 0.4)
- [0.4, 0.6)
- [0.6, 0.8)
- [0.8, 1.0]

Change in gene expression with increased urate

0.2

0.0

−0.2

Figure 6

Table 1: Genes implicated as causal via identification of missense variants with high probability of driving the urate association signal. Genes are included if they contain a missense variant with posterior probability of association of >50% or mapping into a small credible set (≤5 variants).

Gene	SNP	#SNPs in set	SNP PP	consequence	CADD	DHS	Gout p-value (EA)	Brief summary of literature and gene function
ABCG2	rs2231142	4	0.41	p.Gln141Lys (NP_004818.2)	18.2	ENCODE epithelial	1.21E-290	Encodes a xenobiotic and high-capacity urate membrane transporter expressed in kidney, liver and gut. Causal variants have been reported for gout susceptibility (#138900) and the Junior Jr(a-) blood group phenotype (#614490). The locus was first identified in association with serum urate through GWAS (PMID:18834626) and confirmed in many studies since. The common causal variant Q141K has been experimentally confirmed (PMID:19506252) as a partial loss of function.
UNC5CL	rs742493	4	0.95	p.Arg432Gly (NP_775832.2) (within Death domain)	21.0	ENCODE epithelial	2.73E-01	Encodes for the death-domain-containing Unc-5 Family C-Terminal-Like membrane-bound protein. Suggested as a candidate gene for mucosal diseases, with a role in epithelial inflammation and immunity (PMID:22158417). Experiments using human HEK293 cells showed that UNC5CL can transduce pro-inflammatory programs via activation of NF-□B, with the 432Gly variant less potent to do so than the 432Arg one (PMID:22158417).
HNF1A	rs1800574	2	0.92	p.Ala98Val (NP_000536.5)	23.4		1.83E-02	Encodes a transcription factor with strong expression in liver, guts and kidney. Rare mutations cause autosomal-dominant MODY type III (#600496). Locus found in GWAS of T2DM (PMID:22325160) and blood urea nitrogen (PMID:29403010). Together with HNF4-alpha, it was first recognized as master regulator of hepatocyte and islet transcription. Knockout mice show proximal tubular dysfunction (Fanconi syndrome). HNF1A enhanced promoter activity of PDZK1, URAT1, NPT4 and OAT4 in human renal proximal tubule cell-based assays (PMID:28724612), supporting a role in the coordinated expression of components of the urate "transportosome".
HNF4A	rs1800961	1	1.00	p.Thr139Ile (NP_000448.3)	24.7	ENCODE pancreas	7.43E-03	Encodes another nuclear receptor and transcription factor that controls expression of many genes, including *HNF1A* and other overlapping target genes. Rare mutations cause autosomal-dominant MODY type I (#125850) and autosomal-dominant renal Fanconi syndrome 4 (# 616026). Shown to regulate expression of SLC2A9 and other members of the urate "transportosome" in cell-based assays (PMID 25209865, PMID:30124855). The GWAS locus has been reported for multiple cardio-metabolic traits and T2DM (PMID:21874001).
CPS1	rs1047891	84	0.84	p.Thr1412Asn (NP_001116105.1)	22.1		5.66E-02	Encodes mitochondrial carbamoyl phosphate synthetase I, which catalyzes the first committed step of the urea cycle by synthesizing carbamoyl phosphate from ammonia, bicarbonate, and 2 molecules of ATP. Rare mutations cause autosomal-recessive carbamoylphosphate synthetase I deficiency (#237300). In addition to hyperammonemia, this disease features increased synthesis of glutamine, a precursor of purines. Elevated uric acid excretion has been reported in patients with hyperammonemia (PMID:6771064). GWAS locus for eGFR (PMID:26831199), homocysteine (PMID:23824729), urinary glycine concentrations (PMID: 26352407).
GCKR	rs1260326	2	0.67	p.Leu446Pro (NP_001477.2)	0.1	ENCODE kidney	4.09E-41	Encodes a regulatory protein prominently expressed in the liver that inhibits glucokinase. Identified in previous GWAS of urate (PMID:23263486) and multiple other cardio-metabolic traits. The 446L protein was shown to be less activated than 446Pro by physiological concentrations of fructose-6-phosphate, leading to reduced glucokinase inhibitory ability (PMID:19643913).

Abbreviation: PP, posterior probability; DHS, DNase-I hypersensitivity site; CADD, Combined Annotation Dependent Depletion phred score; EA, European ancestry.